

NBER WORKING PAPER SERIES

GAMING IN AIR POLLUTION DATA? LESSONS FROM CHINA

Yuyu Chen  
Ginger Zhe Jin  
Naresh Kumar  
Guang Shi

Working Paper 18729  
<http://www.nber.org/papers/w18729>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
January 2013

This paper is published in *The B.E. Journal of Economic Analysis & Policy, Advances*, Volume 13, Issue 3, 2012. We thank Editor Eric Zitzewitz, two anonymous referees, Douglas Almond, Hanming Fang, Erica Li, and the participants at the 2011 NBER China Group meeting, the 2012 International Industrial Organization Conference, and the 2012 NBER conference on public finance issues in China for constructive comments. Financial support from Peking University is gratefully acknowledged. All errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Yuyu Chen, Ginger Zhe Jin, Naresh Kumar, and Guang Shi. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Gaming in Air Pollution Data? Lessons from China  
Yuyu Chen, Ginger Zhe Jin, Naresh Kumar, and Guang Shi  
NBER Working Paper No. 18729  
January 2013, Revised January 2013  
JEL No. D8,H7,I18,L3,L5

### **ABSTRACT**

Protecting the environment during economic growth is a challenge facing every country. This paper focuses on two regulatory measures that China has adopted to incentivize air quality improvement: publishing a daily air pollution index (API) for major cities since 2000 and linking the API to performance evaluations of local governments. In particular, China defines a day with an API at or below 100 as a blue sky day. Starting in 2003, a city with at least 80% blue sky days in a calendar year (among other criteria) qualified for the “national environmental protection model city” award. This cutoff was increased to 85% in 2007.

Using officially reported API data from 37 large cities during 2000-2009, we find a significant discontinuity at the threshold of 100 and this discontinuity is of a greater magnitude after 2003. Moreover, we find that the model cities were less likely to report API right above 100 when they were close to the targeted blue sky days in the fourth quarter of the year when or before they won the model city award. That being said, we also find significant correlation of API with two alternative measures of air pollution – namely visibility as reported by the China Meteorological Administration (CMA) and Aerosol Optical Depth (AOD), corrected for meteorological conditions, from NASA satellites. The discontinuity around 100 suggests that count of blue sky days could have been subject to data manipulation; nevertheless, API does contain useful information about air pollution.

Yuyu Chen  
Applied Economics Department  
Guanghua School of Management  
Peking University  
Beijing, 100871  
Chenyuyu@gsm.pku.edu.cn

Ginger Zhe Jin  
University of Maryland  
Department of Economics  
3115F Tydings Hall  
College Park, MD 20742-7211  
and NBER  
jin@econ.umd.edu

Naresh Kumar  
Department of Epidemiology and Public Health  
Miller School of Medicine  
University of Miami  
naresh-kumar@uiowa.edu

Guang Shi  
Rm 219  
Development Research Center  
of the State Council of China  
Chaonei Street, Dongcheng District  
Beijing, China, 100010  
shiguangpku@gmail.com

## 1. Introduction

Protecting the environment during economic growth is a challenge facing every country. Because private sectors may not fully internalize the environmental consequences of their activity, environmental protectionists have called for government regulation. Existing research has examined the effect of environmental regulations on firm behavior<sup>1</sup>, environmental measures<sup>2</sup>, and health outcomes<sup>3</sup>, but few studies look at the policies that motivate *local governments* to reduce pollution.<sup>4</sup> Nevertheless, a great number of international treaties have been established, with the assumption that each targeted government, facing the incentives specified in an international treaty, can effectively reduce pollution in or near its territory. Unfortunately, any policy that motivates local governments to reduce pollution can also motivate them to report better outcomes on paper, especially if it is more costly to make actual improvements, if gaming is difficult to detect, and if enforcement relies on self-reporting due to a lack of disaggregated objective data.

China provides a unique opportunity to study local government incentives in environmental protection. While China has enjoyed steady GDP growth for the past 30 years, 16 of the world's top 20 most polluted cities were located in China as of 2007.<sup>5</sup> Given the regional decentralized authoritarian (RDA) regime in China, Xu (2011) argues that local government officials have an incentive to sacrifice environmental protection in order to boost local GDP growth. This is because local government leaders are appointed by the central government based on local performance, and GDP growth is easier to measure than environmental conditions. To be fair, the central government is aware of the problem and has

---

<sup>1</sup> Among others, Henderson (1996), Becker and Henderson (2000), Greenstone (2002), and List et al. (2003) examine the effects of environmental regulations on firm entry, exit, and size change in the US. A number of other studies focus on the effect of environmental regulations on trade flow, capital flow, and international pollution havens, for example, Dean, Lovely and Wang (2009), Ederington, Levinson and Minier (2005), Keller and Levinson (2002), Wheeler (2004), and Zeng and Zhao (2009).

<sup>2</sup> For example, Greenstone (2004) studies the impact of the US Clean Air Act on sulfur dioxide, Davis (2008) studies the effect of driving restrictions on air quality in Mexico City, and Chen et al. (2011) study the effect of environmental measures adopted in the name of the 2008 Olympic Games on Beijing's air quality.

<sup>3</sup> For example, Chay and Greenstone (2003), Currie and Neidell (2005), and Currie, Neidell and Schmieder (2009) study the impact of air pollution on infant health and mortality; Chay, Dobkin and Greenstone (2003) examine the effect of the 1970 Clean Air Act on adult mortality.

<sup>4</sup>At the country level, Congleton (1992) and Murdoch and Sandler (1997) show that democratic countries are more likely to support and enforce chlorofluorocarbon emissions control under the Montreal Protocol. See Oats and Portney (2003) for a review of the political economy of environmental policy.

<sup>5</sup><http://www.cbsnews.com/stories/2007/06/06/eveningnews/main2895653.shtml>, citing the World Bank's "[The Little Green Data Book](#)" (May 2007, ISBN 0-8213-6967-9).

incorporated some environmental measures into the performance evaluation criteria. Lessons learned from these environmental incentives are likely to have important implications for crafting other domestic and international policies that target local governments for environmental protection.

This paper focuses on two regulatory measures that China has adopted to incentivize local governments to improve air quality: one is publishing daily air pollution index (API) for 86 cities since 2000; the other is linking the API to performance evaluation of local governments. In particular, China defines a day with an API at or below 100 as a blue sky day. Starting in 2003, a city with at least 80% blue sky days in a calendar year (among other criteria) qualified for the “national environmental protection model city” award. This API cutoff for a model city increased further to 85% in 2007. While these incentive policies were adopted to reduce air pollution, they also provided incentives to game the API data, as the API data are reported by local governments and misreporting is less costly than actual improvement of air quality.

Using the officially reported API data from 37 large cities during 2000-2009, we show that there is a significant discontinuity at the threshold of 100, despite the fact that API is calculated as a city-day average for multiple pollutants across multiple monitoring stations. This discontinuity is of a greater magnitude after 2003, and model cities are less likely to report API right above 100 when they face more pressure to reach the cutoff by the end of the year in which and the year before they won the award. That being said, we also find significant and stable correlation between the API with two alternative measures of air pollution: visibility (reported by the China Meteorological Administration, CMA) and Aerosol Optical Depth (AOD) derived from the NASA satellites. These findings suggest that the count of blue sky days may be subject to data manipulation, but the reported API does contain useful information about air pollution.

We also show that, controlling for weather, day fixed effects, and city-specific factors, there is no statistically significant improvement in the API, visibility, or the AOD immediately before or after a city won the model city award. This implies that model city status is not awarded to acknowledge significant air quality improvement within a city, which could be explained by either the model city policy providing little incentive to improve air quality or the policy encouraging every city to improve on similar scales regardless of model city award status. Both are consistent with the fact that air quality is only a small part of model city award evaluation and the model city award is only one of many incentive tools facing local governments.

The rest of the paper is organized as follows: Section 2 describes data and policies on blue sky days and model city evaluation. Section 3 reviews the literature. Section 4 presents evidence of the discontinuity of the API, the particulate matter inferred from the API ( $PM_{10}$ ), visibility, and the AOD. Section 5

examines the extent to which the pressure to achieve the model-city goal of blue sky days has affected the reported API and PM<sub>10</sub>. Section 6 offers a broader study of whether various measures of air quality have improved immediately before or after a city won the model city award. Section 7 checks the correlations between the API, visibility, and AOD in light of the API discontinuity around the blue sky threshold. Section 8 concludes.

## 2 Data and Background

China has been known for poor air quality since the 1990s. The 1996 national standards on sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), total suspended particles (TSP), and particulate matter with an aerodynamic diameter of 10 microns or smaller (PM<sub>10</sub>) were 2-7 times higher than the standards established by the World Health Organization (UNEP 2009). An amendment in 2000 further weakened the Chinese standards for NO<sub>2</sub> and ozone. Even so, the relatively liberal standards are hard to enforce in China, partly because each local environment protection agency, although a branch of the Ministry of Environmental Protection (MEP), is also part of the local government and thus subject to local governance.

### 2.1 API

Among other environmental protection efforts, the MEP started publishing a daily air pollution index (API) for 86 cities in June 2000. These cities cover most median- and large-size cities of China, including all the provincial-level municipalities and all provincial capitals. For each city, the MEP aggregates the measured intensities of NO<sub>2</sub>, SO<sub>2</sub>, and PM<sub>10</sub> (monitored at sparsely distributed stations with unknown locations in the city) into a daily API ranging from 0 to 500.<sup>6</sup> Specifically, suppose a city has M stations and each station monitors NO<sub>2</sub>, SO<sub>2</sub>, and PM<sub>10</sub> N times each day.<sup>7</sup> The MEP first computes the daily average of all the MxN measures for each pollutant and then translates the daily mean intensity into a pollutant-specific API according to linear spines with cutoff points as defined in Table 1.<sup>8</sup> The overall API is the maximum of all pollutant-specific

---

<sup>6</sup>According to Andrews (2008) and Jiang et al. (2004), API was calculated based on TSP, NO<sub>2</sub>, SO<sub>2</sub>, in 1998-2000. A new policy starting in June 2000 changed API calculation to PM<sub>10</sub>, NO<sub>2</sub>, and SO<sub>2</sub>. MEP monitors the intensity of CO but does not include it in the current API calculation because the calculation formula was set ten years ago and at that time the vehicle volume in China was very low. MEP is considering adding CO and other pollutants for future API. Source: <http://news.163.com/09/0312/11/5470SBA9000120GU.html>

<sup>7</sup> The MEP stipulates the number of monitoring stations according to city population and the size of the established area. For a large city like Beijing, one monitoring station is required for every 25-30 km<sup>2</sup> and the total number of stations must be at least 8.

<sup>8</sup> For example, if the daily mean of PM<sub>10</sub> is 220 µg/m<sup>3</sup>, the corresponding API of PM<sub>10</sub> is (220-150)/(350-150)\*(200-100)+100 = 135.

APIs. If that maximum is above 500, the overall API is capped at 500. The MEP partitions the API into five groups: 0-50, 51-100, 101-200, 201-300, and 301-500, representing “excellent,” “good,” “slightly polluted,” “moderately polluted,” and “heavily polluted” air quality, respectively. The MEP also reports the category of the dominant pollutant(s) if the API is above 50. Over our analysis sample (from 06/06/2000 to 10/31/2009 for 37 cities), 72.9% of API observations reported PM<sub>10</sub> as the main pollutant, 0.35% reported NO<sub>2</sub>, and 6.85% reported SO<sub>2</sub>. The remaining 19.9% have an API below 50.

**Table 1: MEP cutoff points for different levels of API**

API	PollutantDensity( $\mu\text{g}/\text{m}^3$ )			Air quality level	Air Quality condition	Notes of health effects
	PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>			
500	600	2620	940	V	Heavy pollution	Exercise endurance of the healthy people decreases; some will have strong symptoms. Some diseases will appear.
400	500	2100	750			
300	420	1600	565	IV	Moderate pollution	The symptoms of the patients with cardiac and lung diseases will be aggravated remarkably. Healthy people will experience a drop in endurance and increased symptoms.
200	350	250	150	III	Slightly polluted	The symptoms of the susceptible are slightly aggravated, while healthy people will have stimulated symptoms.
100	150	150	100	II	Good	Daily activity will not be affected.
50	50	50	50	I	Excellent	Daily activity will not be affected.

Source: The first four columns are taken from the MEP website. The last three columns are copied from Table 2.2 of UNEP (2009).

Although the API data are disclosed on the MEP website, they are collected and reported by local MEP branches. At the frequency of city-day, it is virtually impossible for the central MEP to verify every number reported by a local branch. To the extent that local MEP officials are subject to local governance, the reliability of the API data may depend on the data collection method (defined by the MEP), as well as local political incentive to report a good number to the central government.

## 2.2 Blue Sky Day

A crude categorization of air pollution refers to a day with an API at or below 100 as a “blue sky” day. Both national and local environmental authorities have used the number of blue sky days in a year as a measure of air pollution. For example, Beijing claims steady air quality improvement because the number of blue sky days increased from 274 in 2008<sup>9</sup> to 285 in 2009<sup>10</sup> and 286 in 2010.<sup>11</sup> However, in our analysis sample, the average API of all blue sky days increased from 65.62 in 2008 to 71.11 in 2009 (up to 10/31/2009). This finding implies that continuous API and binary count of blue sky days can paint different pictures of air quality. Nevertheless, the number of blue sky days is visible in mass media and its improvement is often cited by local governments as a political goal at the beginning of a calendar year and an achievement at the end of the year.<sup>12</sup> The phrase “blue sky day” has also been challenged by a local resident of Beijing, who photographed the sky each day and found that the number of days with a real blue sky was 180 instead of 285 in 2009.<sup>13</sup> While the naked eye and reported data differ in definition, this incident reflects the public attention paid to blue sky days.

## 2.3 Model city policy

As early as 1997, the central government of China started to evaluate whether a city was qualified for a “national environmental protection model city” award based on environmental quality and economic measures. These measures cover air pollution, water quality, noise, percent of green land, management of industrial and residential waste, consumer satisfaction with the environment, municipal institutions that focus on environmental protection, GDP per capita, GDP growth rate, population growth rate, energy consumption, and water consumption.<sup>14</sup> While the API has always been the only measure of air quality in the evaluation criteria, its use in model city evaluation is vague. The ambiguity was reduced over time, as a 2003 regulation specified that a model city must have

---

<sup>9</sup> See news report at <http://news.qq.com/a/20081231/001928.html> (reported on 12/31/2008, accessed on 02/10/2012).

<sup>10</sup> See news report at <http://news.qq.com/a/20091231/001311.html> (reported on 12/31/2009, accessed on 02/10/2012).

<sup>11</sup> See news report at [http://news.xinhuanet.com/fortune/2011-01/01/c\\_13672919.htm](http://news.xinhuanet.com/fortune/2011-01/01/c_13672919.htm) (reported on 1/1/2011, accessed on 02/10/2012).

<sup>12</sup> See news report at <http://news.163.com/11/1107/01/7IyIIFB100014AED.html> (reported on 07/11/2011, accessed on 02/10/2012).

<sup>13</sup> See news report at [http://news.xinhuanet.com/society/2010-10/27/c\\_12704483.htm](http://news.xinhuanet.com/society/2010-10/27/c_12704483.htm) (reported on 10/27/2010, accessed on 02/10/2012).

<sup>14</sup> We could not find any MEP documents that specify the numerical weight of each model-city evaluation measure, nor does the MEP clarify a sufficient condition for the model-city award. Every condition described in the MEP documents appears to be a necessary condition.

over 80% of days in a calendar year with an API below 100 and this cutoff was increased to 85% in 2007.<sup>15</sup>

It is worth noting that participation in the model city evaluation is voluntary. We do not observe which cities applied for the award in a given year, and we assume that every city that satisfies the explicit criteria apply. Upon application, the central government's evaluation committee will visit the city in person and announce the winner(s) afterwards.<sup>16</sup> Among the 37 big cities for which we have complete API and visibility data from 6/5/2000 to 10/31/2009, nine have won the model city award in our sample period: Qingdao (2000), Hangzhou (2001), Changchun (2002), Nanjing (2003), Fuzhou (2004), Shenyang (2004), Nantong (2006), Tianjin (2006) and Guangzhou (2007). Another six cities won the award before the start of our sample: Shenzhen (1997), Dalian (1997), Xiamen (1997), Haikou (1999), Shantou (1999) and Suzhou (1999). Because we do not know the exact timing of award announcement, we assume that the evaluation is conducted based on the data for the year of award and one year before the award.<sup>17</sup>

The model city award is semi-permanent. According to the MEP regulation, a city that has won the model city award is subject to reexamination every three years; if it fails the reexamination, it has two years to correct the problem; if underperformance remains after the correction period, the award will be revoked. In reality, some reexaminations were conducted more than three years after the initial award, and some cities were even exempted from reexamination.<sup>18</sup> To our best knowledge, no model city award has ever been revoked. This suggests that it is more difficult to earn a model city award than it is to keep it. It may also

---

<sup>15</sup> The 1997 regulation was a pilot program that specified the criterion of air quality as "API<100." The 2003 regulation clarified the air quality criterion as the percentage of days with API<100 higher than 80%. The regulation was issued on 11/19/2002 and made effective on 7/1/2003 ([http://www.mep.gov.cn/gkml/zj/bgt/200910/t20091022\\_173806.htm](http://www.mep.gov.cn/gkml/zj/bgt/200910/t20091022_173806.htm)). The 2007 regulation (effective as of 1/1/2007) stipulated that the percentage of days of API<=100 be higher than 85%. Although the 1997 and 2003 regulations specified API<100 instead of API<=100, we believe the actual implementation was always API<=100 because both definitions of blue sky days and the "good" API category use 100 as the upper bound. This assumption is also confirmed in the below discontinuity study and a local MEP branch website in Xiamen.

<sup>16</sup> We could not find any MEP document that provides details on the specific timing of the model city award.

<sup>17</sup> In one particular application, we observe the applicant citing environmental and economic measures in the past two years.

<sup>18</sup> For example, Yangzhou earned the award in 2002 but was re-examined in 2006 ([http://www.mep.gov.cn/gkml/zj/bgth/200910/t20091022\\_174312.htm](http://www.mep.gov.cn/gkml/zj/bgth/200910/t20091022_174312.htm)); Changchun earned the award in 2002, but was re-examined in 2008 ([http://www.mep.gov.cn/gkml/zj/bgth/200910/t20091022\\_174443.htm](http://www.mep.gov.cn/gkml/zj/bgth/200910/t20091022_174443.htm)). According to [http://wfs.mep.gov.cn/mfcs/mfcsxx/gldt/200503/t20050327\\_65797.htm](http://wfs.mep.gov.cn/mfcs/mfcsxx/gldt/200503/t20050327_65797.htm), some model cities can be exempted from re-examination.



reduce incentives for both manipulation and genuine improvements in air quality after winning the model city award, a hypothesis we will examine in Section 6.

## 2.4 Visibility

In addition to the API, we employ two additional proxies for air pollution. The first proxy is visibility, defined as the greatest distance at which an observer with normal eyesight can discern a dark object from the horizontal sky. We obtained daily visibility data from the China Meteorological Administration (CMA), along with local temperature, precipitation, barometric pressure, sunshine, humidity, and wind velocity as reported at 2PM each day at a fixed point in each city. Researchers have shown that API and visibility are negatively correlated (Che et al. 2006, Fan and Li 2008), and visibility is considered to be an important predictor of fine particulates (Ozkaynak et al. 1985, Huang et al. 2009).<sup>19</sup> Like the API, visibility is reported by government officials, but it is not disclosed to the public (we purchased the data from CMA) and not used in the evaluation of government officials, and therefore is subject to fewer gaming incentives.

## 2.5 AOD

The second proxy for air pollution is the daily 10km AOD data (Level 2, collection 5.0) retrieved from Moderate Resolution Imaging Spectroradiometer (MODIS) aboard Terra and Aqua satellites (NASA 2010). The extraction procedure is available elsewhere (Chu et al. 2003; Levy et al. 2007a, 2007b). Imagine that radiation travels from a satellite to the earth's surface. By definition, the AOD captures the amount of radiation absorbed, reflected, and scattered due to the presence of solid and liquid particulates suspended in the atmospheric column (Kaufman, Gobron et al. 2002; Kaufman, Tanre et al. 2002). While the AOD is potentially available everywhere at the satellite crossing time (~10:30am and ~1:30pm of Beijing time), it is sensitive to the point- and time-specific weather and available only for days with less than 10% cloud cover. Despite this fact, researchers have shown that the AOD, corrected for meteorological conditions, can predict air quality (Gupta et al. 2006; Kumar 2010; Kumar et al. 2011). Focusing on Delhi and Kanpur in India and Cleveland in the US, Kumar et al. (2009; 2011) demonstrate how the AOD can be converted to PM<sub>10</sub> estimates. They develop an empirical relationship between *in situ* measurements of PM<sub>10</sub> and the AOD. They conclude that AOD captured 70% of the variations in the PM<sub>10</sub> (monitored on the surface) after controlling for meteorological conditions and

---

<sup>19</sup> Fine particulates (PM<sub>2.5</sub>) are defined as particulates less than 2.5µm in aerodynamic diameter.

seasonality. Since the *in situ* PM<sub>10</sub> data were not available for Chinese cities during our sample period, this paper utilizes the AOD data directly (corrected for meteorological conditions and spatiotemporal structure).

The comparison of API, visibility, and AOD data is far from perfect. Even if the reported dominant pollutant of the API is PM<sub>10</sub>, visibility and AOD data could vary with the composition of particle size. According to Brook, Dann and Burnett (1997), Canadian data suggest that PM<sub>2.5</sub> accounted for 49% of the PM<sub>10</sub>, and PM<sub>10</sub> accounted for 44% of total suspended particles. The composition of suspended particles is likely coarser in developing countries: Kumar and Foster (2009) and Kumar, Chu and Foster (2007) show that PM<sub>2.5</sub> accounts for only 24% of the PM<sub>10</sub> in Delhi, India. Moreover, all particulate matter in the atmosphere could affect the AOD, whereas visibility and the API are more related to particulate matter on the ground. The third difference is due to the mismatch in the spatial resolution of AOD and API. Although we know the centroid latitude and longitude of each AOD data point (which represents ~10km at the satellite crossing path), we do not know the exact location of each API or visibility monitoring station.<sup>20</sup>

## 2.6 Analysis Sample

Conditional on having non-break API and visibility data, our analysis consists of 37 cities, which include major provincial-level municipalities, such as Beijing, Tianjin, Shanghai and Chongqing, as well as 24 provincial capitals.<sup>21</sup>

The API, visibility, and other meteorological data from the CMA are reported by city-day, covering 126,688 observations from 6/5/2000 to 10/31/2009. For the 37 cities in our sample, we retrieved 2,614,734 valid 10km AOD observations from 2/25/2000 to 12/31/2009. Of all the 3,598 calendar days in the sample period, only 49.9% had valid AOD observations due to gaps in the data.<sup>22</sup> On average, we have 39.36 data points of the AOD per city-day.

To control for time-specific meteorological conditions at the observation time of the AOD, we acquired hourly global surface meteorological data from the monitoring stations in and around the selected cities. The details on these data are

---

<sup>20</sup> Andrews (2008) and Viard and Fu (2011) used station-level PM<sub>10</sub> data for Beijing. We are not aware of station-level data available for all the 37 cities in our sample from 2000 to 2009.

<sup>21</sup> Although the MEP reports API for 86 cities and the CMA visibility data cover 69 cities, only 42 cities have API data in 2000 and the visibility data are incomplete for some cities between 1993 and 2009. For an unknown reason, the API data are missing on June 4, 2008 for all cities. So the “non-break” criterion means having valid data every city-day from the beginning to the end of our data sample, ignoring the missing data on June 4, 2008. All the discontinuity analysis in Section 4 is robust on the full sample of 86-city API data.

<sup>22</sup> For 37 cities from February 25, 2000 to December 31, 2009, there are totally 133126 city/day cells. In our data, there are actually 66427 city/day cells with valid AOD observations, which amounts to 49.9%.

available elsewhere (NCDC 2007). These data were collocated with the AOD data within a one-hour time interval of AOD time on a given day. This means we assigned the same value of meteorological conditions (from the closest station) to all AOD values in a given city on a given day. The gaps in meteorological data and AOD data resulted in missing values in 6% of the sample. Therefore, meteorological conditions were imputed for days when AOD was available.<sup>23</sup>

To facilitate the comparison between API, visibility, and AOD data, our AOD analysis focuses on the city-day average of the AOD conditional on AOD availability. This resulted in 63,948 city-day observations of AOD, of which 50,672 city-days reported PM<sub>10</sub> as the dominant pollutant.

### **3 Literature and potential ways to game the API**

We are not the first to question the reliability of API data. Using data for Beijing only, Andrews (2008) expresses concern that Beijing may have manipulated the official API report because (1) Beijing has relocated monitoring stations over time; (2) the 2000 MEP regulation switched one component of API from TSP to PM<sub>10</sub>, and weakened the limits of nitrogen oxides and ozone; and (3) the number of days with an API between 96 and 100 is significantly higher than the number of days with an API between 101 and 105. The article compares reported API with detailed station-level PM<sub>10</sub> data, showing discontinuity of API and PM<sub>10</sub> data, but provides no formal statistical test for the discontinuity. Nor does it control for weather and other factors that may influence API. Guinot (2008) suggests that it is not uncommon to add monitoring stations with economic and urban growth, and the uncertainty in the API metrics may range from 15% to 25% due to measurement errors in pollutant intensities. This casual debate motivates us to examine the discontinuity of API in a more scientific way. Our study also expands from Beijing to 37 big cities, and pays more attention to the incentive of misreport from local governments.

A few other studies have used alternative measures of air pollution in addition to the official API. In February 2009, United Nations published a summary report on the 2008 Beijing Olympic Games, with an entire chapter devoted to air quality (UNEP 2009). This report uses the API and pollution intensity data from the Beijing Environmental Protection Bureau (EPB), plus a brief discussion of coarse resolution AOD data (i.e., 100km instead of the 10km AOD used in our study). The report concludes that Beijing's air quality improved from before to during to shortly after the 2008 Games. Chen et al. (2011)

---

<sup>23</sup>When meteorological variables are missing, we usually miss some but not all of them. Suppose we only miss the meteorological variable K for date t in city c. Conditional on the days when K is available, we regress K on the other meteorological variables and continuous time (days since 2000). We then impute K on date t, using these regression coefficients and other meteorological variables that are available on date t.

investigate the impact of the 2008 Games on air quality using the same API and AOD data as in this paper. After controlling for city-specific attributes and a nationwide trend toward better air, they find that the air-cleaning actions adopted in the name of the 2008 Olympic Games lead to real but temporary improvement in the air quality of Beijing. This result is supported by both API and AOD data, suggesting that the API contains useful information about air pollution. However, this conclusion does not rule out gaming of the API given the imperfect comparison between API and AOD data. Viard and Fu (2011) use both API and station-level  $PM_{10}$  data of Beijing to investigate the impact of traffic restriction on air quality. They find that traffic restriction leads to a 19% decline of API during every-other-day restriction and a 7% decline during one-day-per-week restriction.

Wang et al. (2009) collected their own  $PM_{10}$  and  $PM_{2.5}$  data at Peking University between 7/28/2008 and 10/7/2008. They find a significant correlation between self-measured and published  $PM_{10}$ , but the absolute level of their self-measure is 30% higher. This finding triggers concerns that the official API may be subject to manipulation, but the discrepancy may be attributed to sampling and methodological differences (Tang et al. 2009, Yao et al. 2009, Simorich 2009). Wang et al. (2009) also find that meteorological conditions such as wind, precipitation, and humidity account for 40% of the total variation in  $PM_{10}$ . This is why we need daily meteorological data for every city in our sample.

Any systematic study of gaming needs to ask “why” and “how.” The incentive to improve the reported API is rooted in the unique structure of the Chinese political system. As described in Xu (2011), China is characterized by a combination of political centralization and economic regional decentralization: the central government controls the appointment, promotion, and demotion of local political leaders, while leaving to subnational governments (provinces, municipalities, and counties) the responsibility for initiating and coordinating reforms, providing public services, and making and enforcing laws within their jurisdictions. The central control of personnel is a powerful instrument to induce regional officials to follow the central government’s policies. This so-called regionally decentralized authoritarian (RDA) regime stands in clear distinction to federalism (where governors or mayors are elected from the bottom) and central planning.

Researchers have shown that the central government stipulates performance criteria for local leaders, and these local leaders negotiate narrower and more precisely defined performance contracts with its sub-levels. For example, Tsui and Wang (2004) show that 60 percent of provincial leaders are assigned targets related to economic construction. More generally, work achievement accounts for 60 to 70 percent of the evaluation of regional officials, while political integrity, competence, diligence, and other aspects account for the rest (Edin 2003). Similar personnel control is documented between county

governments and township and village officials (Whiting 2000). Within this structure, every level of government may use absolute and/or relative performance in the political contract for the next level. Maskin, Qian and Xu (2000) provide evidence that officials from relatively better-performing regions have a greater chance of being promoted. Similarly, Chen, Li and Zhou (2005) find that provincial officials' performance relative to the national average and to their immediate predecessors has significant impacts on their promotions. All this evidence suggests that the central personnel control over the local governments is effective and that the model city award policy is likely one of many performance criteria that the central government uses to evaluate local officials.

For gaming of the API to exist, two conditions must hold. First, there needs to be enough noise in the true API that one cannot precisely target a particular number (e.g. the upper bound of blue sky days) via actual improvement. This condition is not difficult to satisfy, given that meteorological conditions such as wind, precipitation, and humidity account for 40% of the total variation in PM<sub>10</sub> (Wang et al. 2009).

The second necessary condition for gaming to exist is that it needs to be difficult to detect. To the extent that the MEP uses the reported API without verification, a local MEP branch could report any number, in theory. However, the reported API will be disclosed to the public, and citizens (including local media) will form their own judgment as to how precise the reported API is relative to their personal experience on that city-day. The recent smog in Beijing demonstrates the high public awareness of air quality and the power of public outcry if the official API is not consistent with personal experience.<sup>24</sup> This suggests that any misreporting cannot stray too far from the truth.

One way to game the system is reporting an API slightly above 100 as slightly below 100. More sophisticated gaming may spread the underreporting if the public cannot distinguish small changes in the API (say, 99 vs. 95). Data manipulation can also be achieved by relocating monitoring stations or computing the aggregated index from a selective sample of existing stations, both of which are difficult to detect because the reported API is not station-specific and *in situ* density of air pollutants was not publicly available in our sample period.<sup>25</sup>

As summarized in Zitzewitz (2012), "forensic economics" relies on several techniques to detect gaming: one is to compare the reported data with

---

<sup>24</sup> The heavy smog in Beijing on December 4-6, 2011 caused extensive flight delays and cancellations. See BBC December 5, 2011 report "China morning round-up: Beijing heavy smog scrutinised" accessed at December 7, 2011 at <http://www.bbc.co.uk/news/world-asia-china-16044098>.

<sup>25</sup> Beijing Municipal Environmental Monitoring Center started to publish pollutant densities by hour and monitoring stations on January 12, 2012. There are not enough historical data to study the reliability of this measure.

other data sources. For example, Fisman and Wei (2004, 2007) and Mishra et al. (2007) compare custom data in both origin and destination countries in order to identify missing imports or missing exports. Zinman and Zitzewitz (2009) compare resort-reported snowfall with official weather data and find that resorts are more likely to over-report snowfalls when they can benefit more from such over-reporting. Snyder and Zidar (2009) compare economists' self-reported publications (in *vitas*) with journals' tables of contents and identify subtle forms of inflation. This method of collation is limited to the availability of other reliable data sources.

A second method to detect gaming is searching for data patterns that are consistent with gaming, for example, bunching around a threshold (Slemrod 1985 and Saez 2010 on income tax, DeGeorge, Patel and Zeckhauser 1999 on earnings management, and Forbes, Lederman and Tombe 2011 on airline delays), patterns that should not exist without cheating (Jacob and Levitt 2003 on school test scores), or a correlation between the reported data and the situations that present strong incentives to game (Michaely and Womack 1999 on stock recommendations, Levitt and Syverson 2008 on real estate sales).

We use both methods. First, we detect discontinuity in the raw data around critical thresholds. Given the clear threshold definition for blue sky days, we expect API density to be lower immediately below 100 than immediately above 100 if gaming exists. Moreover, this discontinuity should be more conspicuous after the central government introduces a quantitative measure of blue sky days in the evaluation of a model city, if local governments care about winning the model city award. With the specified cutoff (80% from 2003 to 2006 and 85% after 2007), we expect cities that are close to the cutoff in the fourth quarter of a year to have an incentive to underreport the API at or below 100. Given the permanence of the model city award, we also expect cities that have already won the model city award to be less eager to game for the award.

Our second approach is to compare reported API with visibility and AOD data. Given the imperfect comparison of the three air pollution proxies, we will investigate their statistical correlations after controlling for other factors that may correlate with them in different ways (e.g. weather). To the extent that the underreporting of the API generates a random number below but close to 100, we expect the correlation between the API and visibility/AOD to be lower when the API is close to 100. However, if underreporting is monotonic to the actual API – for example, report 105 as 100, 104 as 99, and 101 as 96 – gaming does not necessarily predict a lower correlation between the API and visibility/AOD.

## 4 Tests of Discontinuity

This section examines the discontinuity among API, visibility, and AOD data. To the extent that the API cannot be misreported too far from truth, we expect misreporting to lead to less density on the right of 100 than on the left.

### 4.1 Discontinuity of API and underlying $PM_{10}$

The first row of Figure 1 plots the histogram of the API in our whole sample, where bins are defined over the complete API range (0 to 500) with bin width=1 (API is reported in integers). The plot shows likely discontinuity at 50, 100, and 500: the number of observations jumps from 1,408 for 50 to 2,034 for 51, from 1,367 for 99 and 1,005 for 100 to 509 for 101, and from 2 for 499 to 126 for 500.

One possible explanation for this discontinuity is the API definition. As shown in Table 1, the API is a piece-wise linear transformation from the averaged *in situ* measures of the main pollutant on a city-day. As confirmed in simulation, the API definition could generate API discontinuity around each category threshold even if the underlying pollutant density is continuous. To assess this possibility, we focus on the city-days that report  $PM_{10}$  as the main pollutant and infer  $PM_{10}$  density from the reported API. Since the main pollutant is not specified when the API is below 50, our analysis on  $PM_{10}$  focuses on the city-days with the API above 50.

One data issue arises in this process. Because the API is reported in integers and  $PM_{10}$  range (0 to 600) is greater than API range (0 to 500), inferred  $PM_{10}$  has zero density for many  $PM_{10}$  numbers. As shown in the second row of Figure 1, with bin width=1 we observe positive frequency on only 363 out of the 600 potential integers of  $PM_{10}$ . To address this, we group  $PM_{10}$  by a wider bin width of 2, as 2 is the transformation factor from the API to  $PM_{10}$  when the API is below 200 and most of the API data are below 200.<sup>26</sup>

As shown in Figure 1, when bin width=1,  $PM_{10}$  demonstrates apparent discontinuity around 150 (corresponding to API of 100). When bin width is widened to 2, human eyes can still identify discontinuity around  $PM_{10}=150$ .

The second column of Figure 1 presents the Burgstahler and Dichev test (BDT) of discontinuity (Burgstahler and Dichev 1997). In particular, for any bin ( $j$ ) that is not at the boundaries, we compute a BDT statistics by comparing the

---

<sup>26</sup> When API is between 50 and 100,  $PM_{10}=50+(API-50)/(100-50)*(150-50)=50+(API-50)*2$ . When API is between 100 and 200,  $PM_{10}=150+(API-100)/(200-100)*(350-150)=150+(API-100)*2$ . When API is between 200 and 300,  $PM_{10}=350+(API-200)/(300-200)*(420-350)=350+(API-200)*0.7$ . When API is between 300 and 400,  $PM_{10}=420+(API-300)/(400-300)*(500-420)=420+(API-300)*0.8$ . When API is between 400 and 500,  $PM_{10}=500+(API-400)/(500-400)*(600-500)=500+(API-400)$ .

bin's observed relative frequency ( $\hat{p}_j$ ) with the average of frequencies of adjacent bins ( $\hat{p}_{j-1}$  and  $\hat{p}_{j+1}$ ):

$$BDT_j = \frac{\frac{\hat{p}_{j-1} + \hat{p}_{j+1}}{2} - \hat{p}_j}{\sqrt{\text{var}\left(\frac{\hat{p}_{j-1} + \hat{p}_{j+1}}{2} - \hat{p}_j\right)}}$$

where  $n$  is the total number of observations and

$$\begin{aligned} \text{var}\left(\frac{\hat{p}_{j-1} + \hat{p}_{j+1}}{2} - \hat{p}_j\right) &= \frac{1}{n} \hat{p}_j (1 - \hat{p}_j) + \frac{1}{4n} (\hat{p}_{j-1} + \hat{p}_{j+1}) (1 - \hat{p}_{j-1} - \hat{p}_{j+1}) \\ &\quad + \frac{1}{n} \hat{p}_j (\hat{p}_{j-1} + \hat{p}_{j+1}). \end{aligned}$$

According to Burgstahler and Dichev (1997) and Takeuchi (2004),  $BDT_j$  conforms to a standard normal distribution if the true distribution underlying the data is continuous. Obviously, the power of BDT depends on sample size and bin width. Using Monte Carlo simulation, Takeuchi (2004) shows that the test is powerful over moderate sample size ( $n > 500$ ) and is able to detect a small jump if the sample size is large ( $n > 5000$ ). Our sample size is 126,688 for the API and 92,383 for  $PM_{10}$ .<sup>27</sup>

The first graph in the second column of Figure 1 draws BDT against each API value. The two dashed lines correspond to 2.58 and -2.58, the critical values for the 99% confidence in a standard normal distribution. Consistent with the histogram on the left, this BDT graph confirms significant API discontinuity in the neighborhood of 50, 100, and 500. The large BDT at 499 is not surprising given the censoring at 500. The significant BDTs around 50 can be driven by gaming or the piece-wise definition of API. We cannot distinguish the two because the major pollutant is not reported unless the API is greater than 50.

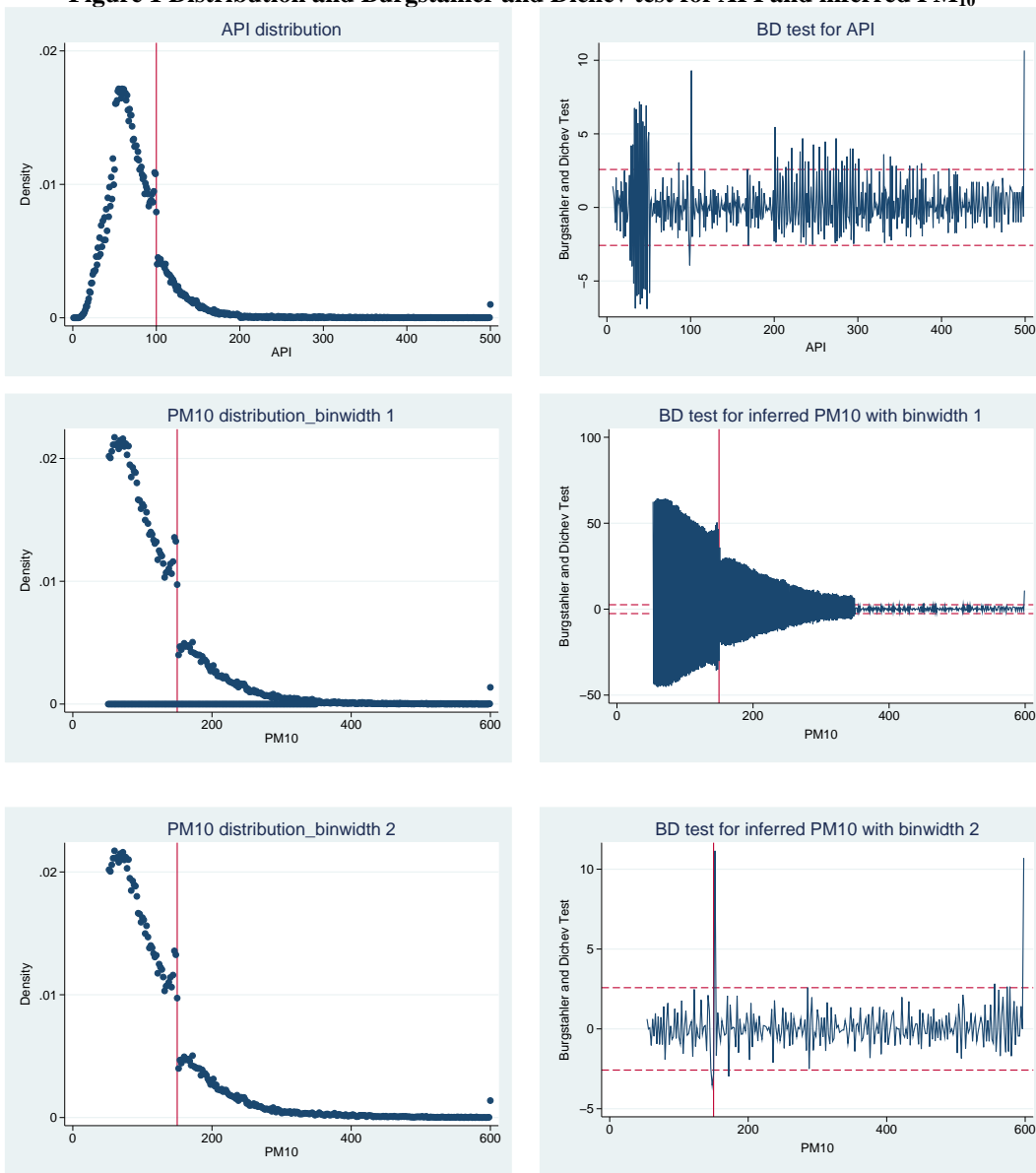
The other two graphs in the second column of Figure 1 present the BDT for each bin of  $PM_{10}$ . With bin width=1 the BDT graph confirms discontinuity at many points (due to gaps in  $PM_{10}$  data), but the discontinuity for  $PM_{10}=150$  still stands out. With bin width=2, the discontinuity at  $PM_{10}=150$  remains significant. This suggests that the discontinuity of API at 100 is not completely driven by its piece-wise definition.

---

<sup>27</sup>One potential caveat of the BDT is that its value is proportional to the square root of sample size. When sample size is very large (e.g. 2.6 million observations for our point-specific AOD data) and the value of each raw data point is limited to a small number of decimal points, each computed BDT can exceed the critical value even if the underlying distribution is continuous. This is partly why we choose to focus on city-day average of the AOD rather than point-specific AOD. While this choice may create some smoothness in the AOD, it is arguably a better comparison with the API, not only because they are both at the city-day level but also because the API by construction is an average of station- and time-specific data on a city-day.



**Figure 1 Distribution and Burgstahler and Dichev test for API and inferred PM<sub>10</sub>**



**Figure 2: Burgstahler and Dichev test for API and PM<sub>10</sub> discontinuity before and after 2003**

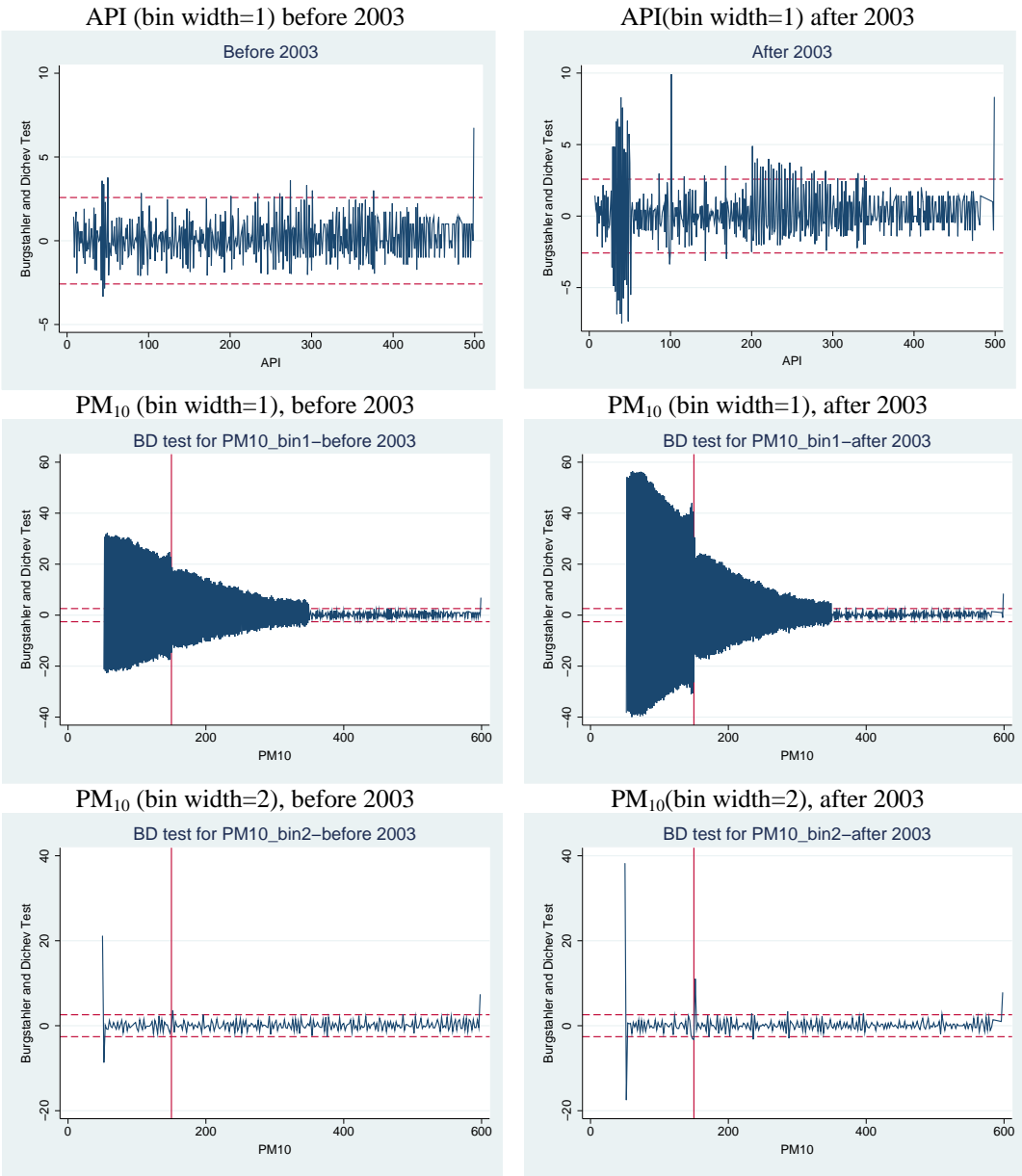


Figure 2 presents the BDT statistics of API,  $PM_{10}$  (bin-width=1) and  $PM_{10}$  (bin-width=2) before and after 2003 separately. The discontinuity around API=100 and  $PM_{10}$ =150 is more significant after 2003, consistent with the introduction of blue sky targets in 2003. An alternative explanation is that there are more data after 2003, and the BDT increases with sample size by definition.

The BDT statistics use only bins adjacent to the study bin. A more general test introduced by McCarry (2008) employs all the data to the left and right of a potential break point and smoothes the histogram by running local linear regressions on the two sides separately. If there is no discontinuity at the break point, the two predicted densities at the break point should be close to each other. This yields a discontinuity estimate (log difference in the two predicted densities at the break point) and the corresponding standard error and t statistics.

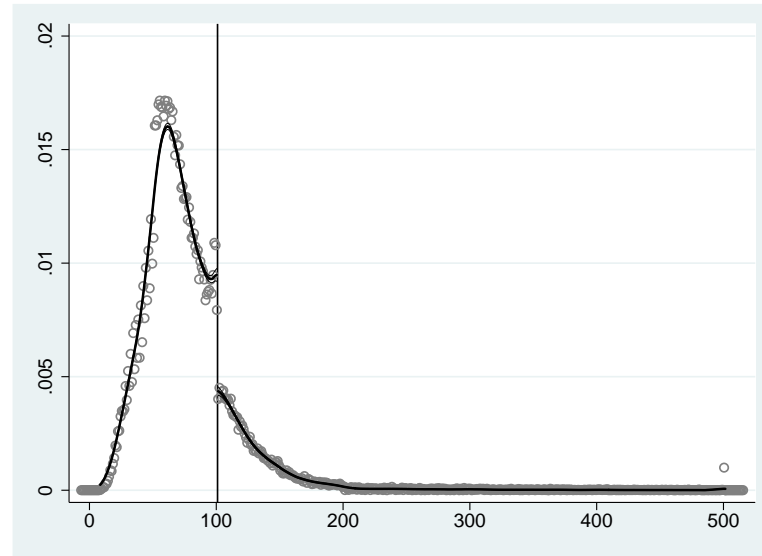
Figure 3A presents the API histogram with smoothed densities to the left and right of 100. Following McCarry (2008), we set bin size as 1 and bandwidth as 15. Below Figure 3A, we report the discontinuity estimate, standard error, and t-test for the full sample of the API and the subsamples of 2000-2002, 2003-2006, 2007-2009, model cities, and non-model cities. By model cities, we mean all the ( $\geq y - 1$ ) observations of a city if that city won the model city award in year  $y$ . Within model cities, we further distinguish model cities in the years preparing for and receiving the award ( $y - 1$  &  $y$ ) versus years after winning the award ( $> y$ ).

Not only does the McCarry test confirm the API discontinuity at 100; it shows that the discontinuity estimate more than doubled from 2000-2002 to 2003-2006 and only declined slightly after 2007. Similarly, the discontinuity estimate for model cities almost triples that of non-model cities. These patterns support the argument that local governments have more incentives to underreport the API since the central government began emphasizing the number of blue sky days in model city evaluation. Note that within model cities, the discontinuity estimate is similar in the years before and after winning the award.

Figure 3B repeats the exercise for inferred  $PM_{10}$ . Given the data gaps in  $PM_{10}$ , we define the x-axis as  $PM_{10}/2$  and conduct the McCarry test with bin size=1 and bandwidth=10. Both the graph and the discontinuity estimates show significant discontinuity around  $PM_{10}$ =150 after 2003, and such discontinuity is more conspicuous for model cities. This pattern is consistent with gaming around  $PM_{10}$ =150 and thus API=100.

Is it possible that the API discontinuity around 100 is driven by local governments targeting 100 in real air quality? If the answer is yes, to the extent that visibility and the AOD are correlated with the API, we shall observe some discontinuity for these two variables as well, especially in the days where the API is not far from 100. Conversely, if the API discontinuity at 100 is driven by gaming, visibility and the AOD should not demonstrate any discontinuity because they are used in neither mass media nor model city evaluation.

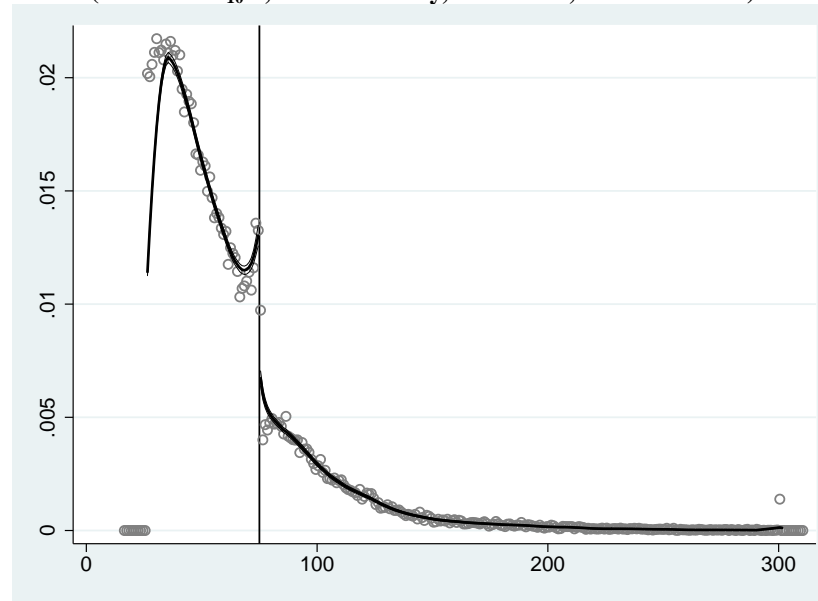
**Figure 3A: McCarry test of the API and discontinuity around 100**  
(X axis: API, Y axis: density, bin size=1, bandwidth=15)



	Full sample	2000-2002	2003-2006	2007-2009	Model cities	Model cities 0-1 years before award	Model cities after award	Non-model cities
Discontinuity estimate	-0.776	-0.409	-0.958	-0.889	-1.568	-1.497	-1.589	-0.536
Standard error	0.029	0.052	0.044	0.060	0.068	0.139	0.078	0.033
T-statistics	-26.744***	-7.901***	-21.787***	-14.941**	-23.005***	-10.752***	-20.335***	-16.358***

\*\*\*  $p < 0.01$ . If a city won the model city award in year  $y$ , the sample of “model cities” includes its daily observations in years in or later than  $y-1$ , the sample of “model cities 0-1 years before award” includes its daily observations in years  $y$  and  $y-1$ , and the sample of “model cities after award” includes its daily observations in years from  $y+1$  and on. The sample of “non model cities” includes every observation that is not in the sample of “model cities.” According to McCarry (2008), the discontinuity estimate represents the log difference in height between the two sides of the discontinuity.

**Figure3B McCarry test for discontinuity of inferred PM<sub>10</sub> around PM<sub>10</sub>=150**  
 (X axis: PM<sub>10</sub>/2, Y axis: density, bin size=1, bandwidth=10)



	Full sample	2000-2002	2003-2006	2007-2009	Model cities	Model cities 0-1 years before award	Model cities after award	Non-model cities
Discontinuity estimate	-.638	-.518	-.692	-.658	-.854	-.865	-.906	-.537
Standard error	.034	.065	.049	.066	.061	.130	.080	.041
T-statistics	-18.962***	-7.925***	-14.184***	-9.948***	-14.059***	-6.655***	-11.305***	-13.224***

\*\*\* p<0.01. If a city won the model city award at year y, the sample of “model cities” includes its daily observations in years at or later than y-1, the sample of “model cities 0-1 years before award” includes its daily observations in years y and y-1, and the sample of “model cities after award” includes its daily observations in years from y+1 and on. The sample of “non model cities” includes every observation that is not in the sample of “model cities.”

Figures 4 and 5 present the BDT for visibility and the AOD. Contrary to our expectation, visibility demonstrates significant discontinuity at integers, especially multiples of five. This phenomenon is easy to explain because visibility is based on human visual observation and manual reporting. The bottom two panels of Figure 4 separate the BDT of visibility by ranges of API (0-40, 40-80, 80-120, 120-500). If the API discontinuity at 100 reflects real air quality and visibility is a valid proxy of air quality, we may observe more discontinuity of visibility when the API is between 80 and 120, as compared to other ranges of the API. Figure 4 does not support this conjecture. As for the city-day average of the AOD, Figure 5 shows no obvious discontinuity at any particular value, which is further confirmed by ranges of the API.<sup>28</sup> These patterns are consistent with the lack of gaming incentives in the AOD data.

Above all, the observed API distribution reveals a significant discontinuity at 100 and such discontinuity is not completely driven by the piece-wise definition of the API. The distributions of the API and inferred PM<sub>10</sub> are consistent with gaming in response to the definition of blue sky days and the model city policy.

## 5. Regression results on model city incentives

This section uses regressions to detect whether cities respond to the targeted number of blue sky days in model city evaluation. Because the pressure for city  $c$  to manipulate the API of day  $d$  in year  $y$  depends on the target and the previously achieved number of blue sky days, we define

$$Pressure_{cya} = \frac{\text{target \# of blue sky days} - \text{realized blue sky days}}{\text{\# of days in the calendar year} - \text{day of year}}$$

For example, if a city realized 16 blue sky days in January 2003, then its pressure on February 1 is  $(365 \cdot 80\% - 16) / (365 - 32) = 0.829$ . If February 1 is not a blue sky day, then its pressure on February 2 is  $(365 \cdot 80\% - 16) / (365 - 33) = 0.831$ .

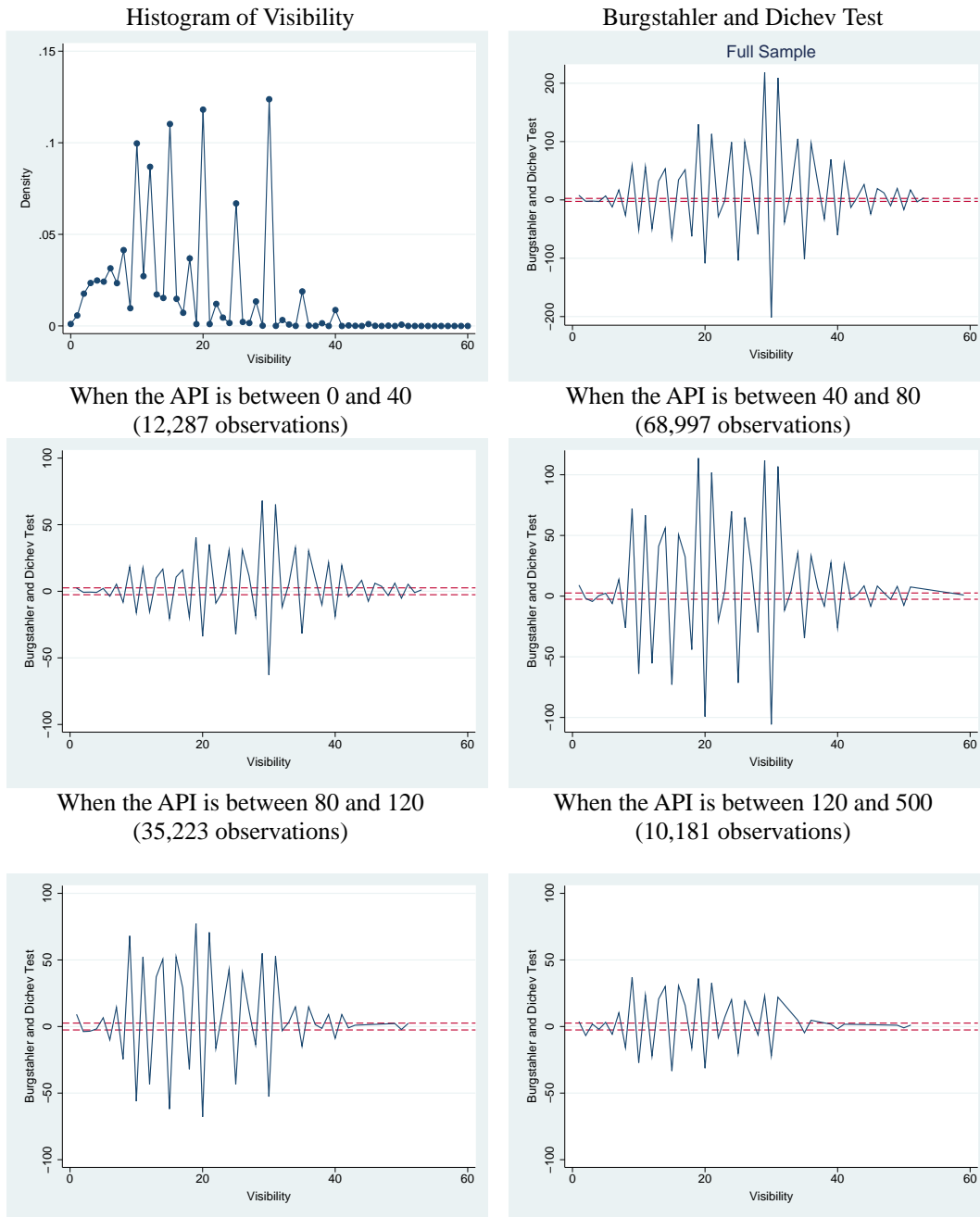
Several adjustments are necessary. First, *Pressure* is coded 0 before 2003, because the blue sky day target was not effective until 2003. Second, if the targeted number of blue sky days is achieved before the end of the year, then *Pressure* is coded 0 after the achievement. Third, *Pressure* is coded 0 if the above formula yields a value larger than 1, which implies that it is impossible to achieve the

---

<sup>28</sup>In unreported graphs, we plot the BDT against a fine grid of point-specific AOD. Probably due to the very large sample size (2.6 million) and the limited decimal points in the AOD data, we have the BDT exceeding the critical value everywhere and across all ranges of the API. Even if we take the BDT values literally, it implies discontinuity everywhere, which is inconsistent with the special discontinuity of API around 100.

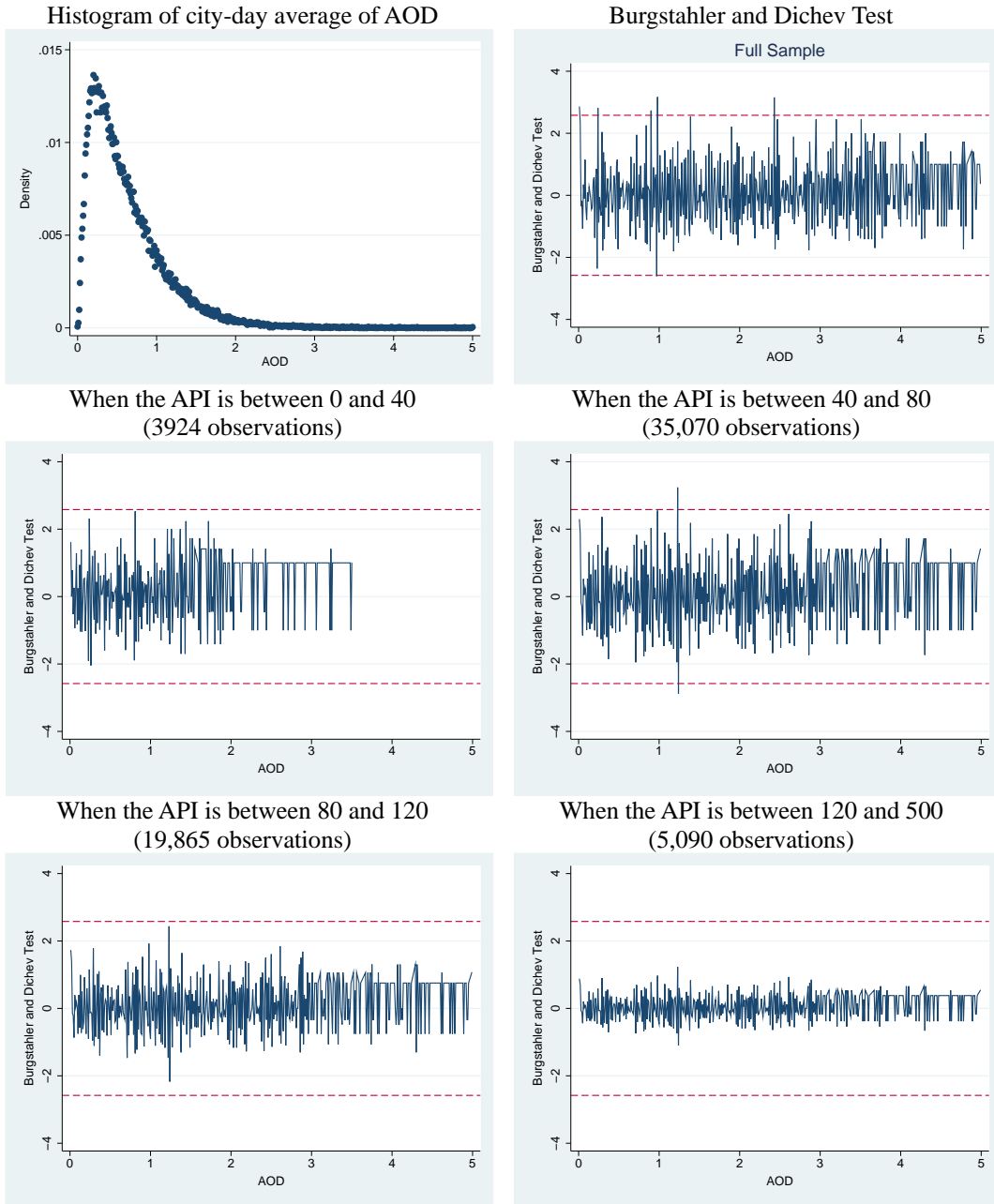
**Figure 4 Histogram and Burgstahler and Dichev Test of Discontinuity for Visibility**

Visibility ranges from 0 to 60 kilometers. Bin width = 1 kilometer.



**Figure 5: Histogram and Burgstahler and Dichev Test of Discontinuity  
for city-day average of AOD**

Bin width = 0.01.





target.<sup>29</sup> Fourth, the above formula is not well defined for the last day of a calendar year. We code *Pressure* 0 for the last day of year, because it should be zero if the target has been realized before the last day. If the target has not been met, there is no way to meet it unless the city is only one day short of the target, which is rare in our data.<sup>30</sup> Above all, two types of variations help identify the effect of *Pressure*: one is the comparison between the days when gaming could help (*Pressure* > 0) and the days when gaming is not necessary or unhelpful; the other is different magnitudes of pressure conditional on *Pressure* > 0.

To capture potential gaming incentives of model city, we define:

$Model_{cy} = 1$  if city *c* is announced as a model city in year *y*;

$Modellag_{cy} = 1$  if city *c* has been announced a model city before year *y*;

$Modelahead_{cy} = 1$  if city *c* is announced as a model city in year *y* + 1.

Given the API discontinuity around 100, we want to understand whether cities systematically underreport an above-100 API number as below 100. To the extent that underreporting must be subtle to avoid public attention, we expect gaming to lead to a lower probability to report right above 100 (defined as between 101 and 105, inclusive) and a higher probability to report right below 100 (defined as between 96 and 100, inclusive). Alternatively, if model cities won the award because of overall improvement in air quality, the whole API distribution should shift to the left. Because the blue sky threshold is on the right tail of the API distribution (Figure 1), real improvement should imply lower API density in both [96,100] and [101,105].

At first glance, it seems straightforward to regress a dummy of whether the reported API falls into [96,100] or [101,105] on *Pressure*, *Model*, *Modellag*, *Modelahead* and their interactions for the full sample. This regression is likely to generate bias because *Pressure* is defined by the API on previous days of the same year. Not only does this present a classical econometrics problem with a lagged dependent variable on the right hand side; it is also possible that local MEP branches (or city governments) engage in sophisticated dynamic programming throughout the year in order to meet the targeted number of blue sky days. Since officials are likely to know more than we do about the benefits and costs of meeting the target, it is difficult for us to capture this dynamic behavior in an explicit model.

---

<sup>29</sup> Results change little if we add a separate dummy to control for the cases of “impossible to reach the target.”

<sup>30</sup>This happens only once in our data, for Yinchuan in 2003, and its last day API reading is 101 in 2003. This city did not win the model city award until 2011.

To address this problem, we focus on the last quarter of each year and take *Pressure* as of September 30 as a predetermined variable. We run the regression:

$$\begin{aligned}
Y_{cyd} = & \alpha_c + \alpha_d + \theta_c \cdot d + \beta_x X_{cyd} + \beta_1 Pressure0930_{cyd} + \beta_2 Model_{cy} \\
& + \beta_3 Modella_{cy} + \beta_4 Modelahead_{cy} \\
& + \gamma_1 Pressure0930_{cyd} \cdot Model_{cy} + \gamma_2 Pressure0930_{cyd} \cdot Modella_{cy} \\
& + \gamma_3 Pressure0930_{cyd} \cdot Modelahead_{cy} + \varepsilon_{1cyd}
\end{aligned}$$

where  $Y_{cyd}$  is the dummies of  $96 \leq API \leq 100$ ,  $101 \leq API \leq 105$ ,  $140 \leq PM_{10} \leq 150$ , or  $151 \leq PM_{10} \leq 160$ . The choice of neighborhood range is arbitrary, but we ran the same regressions by alternative ranges (+/-10, +/-8, +/-3 for the API and +/- 20, +/- 16, +/- 6 for  $PM_{10}$ ) and obtained similar results.<sup>31</sup>  $\alpha_c$  are city fixed effects,  $\alpha_d$  are date fixed effects,  $\theta_c$  are city-specific time trends<sup>32</sup>, and X are control variables including city-day weather<sup>33</sup> and socioeconomic indicators such as GDP growth rate, GDP per capita, industrial production, population, energy consumption, number of private vehicles, and a dummy for regular heating season if heating is provided by the city.

Our main interest is the two-way interaction between *Pressure0930* and the status of model city. If a city won the model city award by gaming the API, gaming should be more apparent in the year of or the year immediately before the award, and when it is subject to a greater pressure to reach the target. Errors are clustered by city.

Linear probability<sup>34</sup> results are reported in Table 2. The first two columns focus on the API in the full sample; the last two columns focus on inferred  $PM_{10}$  when  $PM_{10}$  is the dominant pollutant. Coefficients on *Pressure0930 \* Model* and *Pressure0930 \* Modelahead* suggest that the higher the pressure to reach the target, the more likely it is for a city that is about to win the award to report an API/ $PM_{10}$  right below the blue sky threshold. Similarly, the higher the pressure, the less likely the city is to report an API/ $PM_{10}$  right above the threshold. All these effects are relative to two or more years before winning the award. In comparison, in the years after winning the award, neither API nor TSP regressions

---

<sup>31</sup>The signs of key coefficients remain the same as before, but those in the smaller bands (+/-3 for API and +/-6 for  $PM_{10}$ ) are less significant due to fewer data frequencies in the small bands.

<sup>32</sup>In the reported table, we use a city-specific linear trend. Results are similar when we use a quadratic or cubic trend instead.

<sup>33</sup>Weather variables include rainfall, temperature, barometric pressure, sunshine, humidity (if rainfall is zero), wind velocity, and four dummies for wind direction (east, south, west, and north).

<sup>34</sup> We use a linear probability model because every regression includes a large number of date fixed effects.

**Table 2: Regression results on pressure to meet the target of model city**

**(observation = city-day)**

Linear probability model, observations from the 4<sup>th</sup> quarter only.

$$\text{Pressure0930} = \frac{\text{target \# of blue sky days} - \text{realized blue sky days}}{\text{\# of days in the calendar year} - \text{day of year}} \text{as of 9/30 for 2003 and after.}$$

VARIABLES	Full sample		Sample with PM <sub>10</sub> as dominant pollutant	
	(1) API in [96,100]	(2) API in [101,105]	(1) PM10 in [140,150]	(2) PM10 in [151,160]
pressure0930_model	0.088** (0.033)	-0.047** (0.018)	0.124** (0.049)	-0.060*** (0.019)
pressure0930_modellag	0.027 (0.020)	-0.017 (0.018)	0.019 (0.020)	-0.022 (0.018)
pressure0930_modelhead	0.127*** (0.028)	-0.016 (0.026)	0.054 (0.035)	-0.044** (0.021)
pressure0930	-0.020 (0.019)	0.014 (0.013)	-0.026 (0.020)	0.016 (0.012)
Model	-0.008 (0.027)	-0.004 (0.012)	-0.011 (0.029)	0.001 (0.014)
Modellag	0.011 (0.031)	-0.008 (0.014)	0.027 (0.034)	-0.005 (0.015)
Modelhead	-0.035 (0.023)	0.003 (0.011)	-0.006 (0.023)	0.009 (0.012)
Weather	Y	Y	Y	Y
Date FE	Y	Y	Y	Y
City FE	Y	Y	Y	Y
City-specific trend	Y	Y	Y	Y
Socioeconomic	Y	Y	Y	Y
Observations	31,688	31,688	24,288	24,288
R-squared	0.065	0.067	0.069	0.065

Robust standard errors in parentheses.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Socioeconomic factors include GDP growth rate, GDP per capita, industrial production, population, energy consumption and vehicle number by city and year, as well as the dummy for regular heating season. Weather variables include rainfall, temperature, barometric pressure, sunshine, humidity (if rainfall is zero), wind velocity, and four dummies for wind direction (east, south, west, and north).

suggest significant changes in the neighborhood of the blue sky threshold. This finding is consistent with the semi-permanent nature of the model city award.

The  $PM_{10}$  coefficients suggest that, compared to a city facing 10% pressure, a city facing 60% pressure<sup>35</sup> as of September 30 in the years of winning the award is 6.2 percentage points more likely to report a  $PM_{10}$  right below the blue sky threshold and 3.0 percentage points less likely to report a  $PM_{10}$  right above the threshold. In the years immediately before winning the award, the corresponding numbers are 2.7 percentage points<sup>36</sup> more likely to report a  $PM_{10}$  right below the blue sky threshold and 2.2 percentage points less likely to report a  $PM_{10}$  right above the threshold. These magnitudes are large compared to the 5.88% likelihood of observing a  $PM_{10}$  in [140,150] and 2.28% likelihood of observing a  $PM_{10}$  in [151,160] in the raw data.

For robustness, we rerun the regressions using October 31 instead of September 30 as the cutoff date. This makes the regression sample smaller. Coefficients on the two-way interactions between *Pressure* and model city status are similar to those reported in Table 2, some with larger standard errors.

## 6. A broader evaluation of model city policy

Given the permanency of model city award and the evidence of gaming API around 100, a remaining question is whether model city policies are effective at motivating local governments to engage in an overall improvement of air quality. Specifically, we have two predictions: first, if the model city award is granted to acknowledge air quality improvement of a city, model cities should have more air quality improvement right before winning the award than at least two years before the award. Second, the lax re-examination policy implies that model cities could reduce efforts to improve air quality after winning the award.

Tables 3A and 3B summarize the API and  $PM_{10}$  by city type and city status in terms of model city award. In particular, we distinguish three city types: cities that did not win any model city award by 2010 (total 22 cities), cities that won the model city award in our sample period of 2000-2009 (total 9 cities), and cities that won the award in or before 1999 (total 6 cities). For cities that won the award during our sample period, we further distinguish their observations for at least 2 years before the award, 0-1 years before the award, and all years after the award. As expected, cities that won the award earlier tend to have a lower average API, lower  $PM_{10}$ , and higher counts of blue sky days. Similarly, for the cities that won the award during our sample, API and blue sky days improve over time.

Table 3A also shows that the probability of the API in [96,100] increases during the award-winning years but decreases afterwards. Given the fact that the

---

<sup>35</sup> Among 369 city-year observations, 135 have  $pressure0930=0$ . Conditional positive  $pressure0930$ , the mean of  $pressure0930$  is 0.54 and standard deviation is 0.20.

<sup>36</sup> Of the four numbers, all but 2.7 are significant from zero with 95% confidence.

**Table 3A: Summary of API before, during and after a city won the Model City Award (obs=city-day)**

	City type		
	Did not win model city award before 2010	Won model city award between 2000 and 2009	Won model city award at or before 1999
Number of cities per type	22	9	6
API			
>= 2 years before winning the award	80.21 (42.34)	82.09 (39.37)	
0-1 years before winning the award		77.52 (32.28)	
After winning the award		74.34 (28.68)	56.67 (26.37)
Blue sky day? (=1 if API<=100)			
>= 2 years before winning the award	80.89% (0.39)	79.47% (0.40)	
0-1 years before winning the award		85.91% (0.35)	
After winning the award		89.18% (0.31)	96.34% (0.19)
API in [96,100]			
>= 2 years before winning the award	4.81% (0.21)	5.91% (0.24)	
0-1 years before winning the award		7.21% (0.26)	
After winning the award		6.60% (0.25)	2.00% (0.14)
API in [101,105]			
>= 2 years before winning the award	2.77% (0.16)	2.50% (0.16)	
0-1 years before winning the award		1.80% (0.13)	
After winning the award		1.34% (0.12)	0.48% (0.07)

**Table 3B: Summary of PM<sub>10</sub> before, during and after a city won the Model City Award (obs=city-day)**

	City type		
	Did not win model city award before 2010	Won model city award between 2000 and 2009	Won model city award at or before 1999
Number of cities per type	22	9	6
PM <sub>10</sub>			
>= 2 years before winning the award	125.00 (70.86)	126.20 (68.04)	
0-1 years before winning the award		116.96 (56.30)	
After winning the award		110.96 (49.81)	93.23 (42.86)
PM <sub>10</sub> in [140,150]			
>= 2 years before winning the award	6.90% (0.25)	7.31% (0.26)	
0-1 years before winning the award		9.86% (0.30)	
After winning the award		8.68% (0.28)	4.01% (0.20)
PM <sub>10</sub> in [151,160]			
>= 2 years before winning the award	2.83% (0.17)	2.53% (0.16)	
0-1 years before winning the award		1.85% (0.13)	
After winning the award		1.33% (0.11)	0.79% (0.09)

overall density of API declines in the range of 96 to 105, the non-monotonic change is inconsistent with actual air quality improvement over time but consistent with greater incentive to report the API in [96,100] right before winning the award. In comparison, the probability of the API falling in [101,105] declines both during and after the award winning years. This finding could be consistent with gaming or actual improvement.

To test whether the model city award is granted to acknowledge air quality improvement of a city or motivates model cities to reduce air quality protection after receiving the award, we regress each air quality measure on the timing of model city award while controlling for city fixed effects, date fixed effects, weather, and socioeconomic indicators as specified above. Effectively, the default control cities are those that do not win any model city award before the end of our sample period. Compared with them, a city that won the award in our sample should have better air quality right before winning the award and worse air quality after winning the award if air quality improvement targets the model city award. We define

$$\begin{aligned} Prepare_{c_{yd}} &= 1 \text{ if city } c \text{ won the model city award in year } y \text{ or } y+1, \\ After_{c_{yd}} &= 1 \text{ if city } c \text{ won the model city award in or before year } y-1, \end{aligned}$$

and run regressions at the city-day level:

$$\ln(API_{c_{yd}}) = \alpha_{1c} + \alpha_{1d} + \theta_{1c} \cdot d + \gamma_{11} X_{c_{yd}} + \beta_{11} \cdot prepare_{c_{yd}} + \beta_{12} \cdot After_{c_{yd}} + \epsilon_{1c_{yd}}$$

$$\begin{aligned} \ln(PM10_{c_{yd}}) &= \\ \alpha_{2c} + \alpha_{2d} + \theta_{2c} \cdot d + \gamma_{21} \cdot X_{c_{yd}} + \beta_{21} \cdot prepare_{c_{yd}} + \beta_{22} \cdot After_{c_{yd}} + \epsilon_{3c_{yd}} \end{aligned}$$

$$\begin{aligned} 1(API_{c_{yd}} \leq 100) &= \\ \alpha_{3c} + \alpha_{3d} + \theta_{3c} \cdot d + \gamma_{31} \cdot X_{c_{yd}} + \beta_{31} \cdot prepare_{c_{yd}} + \beta_{32} \cdot After_{c_{yd}} + \epsilon_{3c_{yd}} \end{aligned}$$

$$\begin{aligned} \ln(Visibility_{c_{yd}}) &= \\ \alpha_{4c} + \alpha_{4d} + \theta_{4c} \cdot d + \gamma_{41} \cdot X_{c_{yd}} + \beta_{41} \cdot prepare_{c_{yd}} + \beta_{42} \cdot After_{c_{yd}} + \epsilon_{4c_{yd}} \end{aligned}$$

$$\ln(\overline{AOD}_{c_{yd}}) = \alpha_{5c} + \alpha_{5d} + \theta_{5c} \cdot d + \gamma_{51} \cdot X_{c_{yd}} + \beta_{51} \cdot prepare_{c_{yd}} + \beta_{52} \cdot After_{c_{yd}} + \epsilon_{5c_{yd}}$$

Table 4 shows that the air quality in model cities does not significantly improve in the 0-1 years right before winning the model city award, nor does it decrease after winning the award. This finding is consistent across the API, PM<sub>10</sub>, visibility, the AOD and the dummy of blue sky. In light of the significant air quality improvement found in Beijing around the 2008 Olympic Games (Chen et al. 2011), we rerun the regressions without Beijing, without other Olympic-related cities, and without data for 2008 and 2009. The results are similar.

**Table 4: Model City Award and Air Quality Improvement (observation=city-day)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Log (API)	Log (API)	Log (PM <sub>10</sub> )	Log (PM <sub>10</sub> )	Log (visibility)	Log (visibility)	Log (AOD)	Log (AOD)	API <=100	API <=100
0-1 years before winning model city award	-0.040	-0.029	-0.029	-0.019	-0.037	-0.050	0.046	0.134	0.023	0.028
	(0.027)	(0.043)	(0.029)	(0.040)	(0.029)	(0.034)	(0.109)	(0.098)	(0.037)	(0.041)
After winning model city award	-0.037	-0.041	-0.003	-0.004	0.048	0.033	0.008	0.018	0.003	0.016
	(0.035)	(0.058)	(0.045)	(0.058)	(0.058)	(0.067)	(0.075)	(0.110)	(0.046)	(0.049)
Weather	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Date FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
City FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
City-specific trend		Y		Y		Y		Y		Y
Socioeconomic		Y		Y		Y		Y		Y
Observations	126688	126688	92,383	92,383	126684	126684	126706	126706	12668 8	126688
R-square	0.524	0.547	0.449	0.475	0.533	0.540	0.479	0.495	0.292	0.318

Standard errors are clustered by city and in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Socioeconomic factors include GDP growth rate, GDP per capita, industrial production, population, energy consumption and vehicle number by city and year, as well as the dummy for regular heating season. Weather variables include rainfall, temperature, barometric pressure, sunshine, humidity (if rainfall is zero), wind velocity, and four dummies for wind direction (east, south, west, and north).



## 7. Comparison of API/ PM<sub>10</sub> with visibility and AOD

This section examines whether visibility and AOD are less correlated with API when the API are reported to be close to 100. In particular, we define

$$Left = 1 \text{ if API is in } [96,100],$$

$$Right = 1 \text{ if API is in } [101,105],$$

and run the regressions:

$$\begin{aligned} \ln(Visibility_{cyd}) = & \alpha_{1c} + \alpha_{1d} + \theta_{1c} \cdot d + \gamma_1 \cdot X_{cyd} + \beta_{11} \cdot \ln(API_{cyd}) \\ & + \beta_{12} \cdot Left_{cyd} + \beta_{13} \cdot Right_{cyd} + \beta_{14} \cdot \ln(API_{cyd}) \cdot Left_{cyd} \\ & + \beta_{15} \cdot \ln(API_{cyd}) \cdot Right_{cyd} + \epsilon_{1cyd}; \end{aligned}$$

$$\begin{aligned} \ln(\overline{AOD}_{cyd}) = & \alpha_{2c} + \alpha_{2d} + \theta_{2c} \cdot d + \gamma_2 \cdot X_{cyd} + \beta_{21} \cdot \ln(API_{cyd}) \\ & + \beta_{22} \cdot Left_{cyd} + \beta_{23} \cdot Right_{cyd} + \beta_{24} \cdot \ln(API_{cyd}) \cdot Left_{cyd} \\ & + \beta_{25} \cdot \ln(API_{cyd}) \cdot Right_{cyd} + \epsilon_{2cyd}. \end{aligned}$$

As before, control variables  $X_{cyd}$  include weather and socioeconomic indicators. Errors are clustered by city. We present the regression results of visibility in Table 5 and the AOD in Table 6. In both tables, the left panel is for the API using the full sample; the right panel is for PM<sub>10</sub> using the PM<sub>10</sub> dominant sample. In the PM<sub>10</sub> regressions,  $Left$  and  $Right$  are adjusted according to whether PM<sub>10</sub> is in [140,150] or [151,160].

As expected, the API and PM<sub>10</sub> are negatively correlated with visibility and positively correlated with the AOD. In both tables, the coefficients of  $\ln(API) \cdot Left$  and  $\ln(API) \cdot Right$  are not significantly different from zero with more than 95% confidence.  $\ln(API) \cdot Right$  is marginally significant in Table 5 (with a sign opposite to what we would expect under gaming), but its significance disappears if the regression is run on the PM<sub>10</sub> dominant sample only. This suggests that the stronger correlation between visibility and the API when the API is above 100 is probably driven by pollutants other than PM<sub>10</sub>.

Overall, these results are inconsistent with the gaming prediction that the API and PM<sub>10</sub> should be less correlated with visibility and with the AOD when they are close to the blue sky day cutoff. However, these findings do not rule out all types of gaming. It is possible that the extent of underreporting is not random (e.g. shifting down the API by a constant). Even if underreporting is random, it could go much further below the blue sky threshold. It is also possible that visibility and the AOD are more related to fine particles than the API and PM<sub>10</sub>, and this definitional difference dominates gaming underlying the API or PM<sub>10</sub>.

**Table 5: Correlation between API/PM<sub>10</sub> and Visibility (observation =city-day)**

	Full sample				Sample with PM <sub>10</sub> as the dominant pollutant		
	Ln(visibility)	Ln(visibility)	Ln(visibility)		Ln(visibility)	Ln(visibility)	Ln(visibility)
Ln(API)	-0.381*** (0.034)	-0.381*** (0.034)	-0.404*** (0.035)	Ln(PM <sub>10</sub> )	-0.313*** (0.023)	-0.314*** (0.023)	-0.331*** (0.023)
Ln(API)*Left		0.127 (0.359)	-0.009 (0.373)	Ln(PM <sub>10</sub> )*Left		0.073 (0.271)	0.037 (0.280)
Ln(API)*Right		-1.323* (0.720)	-1.166* (0.669)	Ln(PM <sub>10</sub> )*Right		-0.726 (0.536)	-0.635 (0.500)
Left		-0.589 (1.641)	0.031 (1.705)	Left		-0.362 (1.348)	-0.187 (1.393)
Right		6.151* (3.341)	5.415* (3.103)	Right		3.677 (2.710)	3.217 (2.528)
City FE	Y	Y	Y		Y	Y	Y
Date FE	Y	Y	Y		Y	Y	Y
City-specific trend			Y				Y
Socioeconomic			Y				Y
Observations	126684	126684	126684		92,379	92,379	92,379
R-squared	0.561	0.561	0.570		0.573	0.573	0.581

Standard errors are clustered by city. Standard errors in parentheses.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Socioeconomic factors include GDP growth rate, GDP per capita, industrial production, population, energy consumption, and vehicle number by city and year, as well as the dummy for regular heating season. Weather variables include rainfall, temperature, barometric pressure, sunshine, humidity (if rainfall is zero), wind velocity, and four dummies for wind direction (east, south, west, and north).

**Table 6: Correlation between API/PM<sub>10</sub> and city-day average of AOD (observation =city-day)**

	Full sample				Sample with PM <sub>10</sub> as the dominant pollutant		
	Ln(AOD)	Ln(AOD)	Ln(AOD)		Ln(AOD)	Ln(AOD)	Ln(AOD)
Ln(API)	0.415*** (0.042)	0.410*** (0.042)	0.427*** (0.045)	Ln(PM <sub>10</sub> )	0.285*** (0.026)	0.279*** (0.026)	0.185*** (0.020)
Ln(API)*Left		0.417 (0.508)	0.555 (0.496)	Ln(PM <sub>10</sub> )*Left		0.442 (0.367)	0.299 (0.268)
Ln(API)*Right		0.771 (0.907)	0.675 (0.923)	Ln(PM <sub>10</sub> )*Right		0.113 (0.806)	0.371 (0.494)
Left		-1.877 (2.324)	-2.503 (2.269)	Left		-2.158 (1.826)	-1.469 (1.335)
Right		-3.556 (4.203)	-3.110 (4.277)	Right		-0.548 (4.071)	-1.860 (2.493)
City FE	Y	Y	Y		Y	Y	Y
Date FE	Y	Y	Y		Y	Y	Y
City-specific trend			Y				Y
Energy			Y				Y
Socioeconomic			Y				Y
Observations	63948	63948	63948		50,672	50,672	50,672
R-squared	0.638	0.639	0.641		0.660	0.660	0.587

Standard errors are clustered by city. Standard errors in parentheses.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Socioeconomic factors include GDP growth rate, GDP per capita, industrial production, population, energy consumption and vehicle number by city and year, as well as the dummy for regular heating season. Weather variables include rainfall, temperature, barometric pressure, sunshine, humidity (if rainfall is zero), wind velocity, and four dummies for wind direction (east, south, west, and north).

## 8. Conclusion

Overall, this paper focuses on two regulatory measures that China has adopted to incentivize air quality improvement: publishing daily API for major cities since 2000 and linking the count of blue sky days ( $API \leq 100$ ) to the evaluation of the model city award. Using daily API, visibility, and the AOD data from 37 large cities during 2000-2009, we show that the officially reported API has a significant discontinuity at the blue sky threshold. This discontinuity cannot be fully explained by the piece-wise definition of the API and is more pronounced after the introduction of the blue sky target in model city evaluation. These patterns suggest data manipulation around the blue sky threshold.

That being said, we also find significant correlation of the API and two alternative measures of air quality (visibility and the AOD), and such correlations do not change significantly when the API is closely below or above 100. These findings suggest that although count of blue sky days may be subject to data manipulation, the reported API does contain useful information for cross-city and cross-time variations of air pollution.

Evidence regarding the effect of the model city award is less clear. On the one hand, reported API (and inferred  $PM_{10}$ ) around the blue sky threshold tends to be more sensitive to the pressure of reaching the targeted blue sky days when a city was winning the award, and this sensitivity declines after the city won the award. This is consistent with gaming around the threshold. On the other hand, based on all data, there is no statistically significant improvement in the API, inferred  $PM_{10}$ , visibility, or the AOD immediately before or after a city won the model city award once we control for weather, city fixed effects, and date fixed effects.

How then can these mixed findings be explained? We offer several possible explanations. The first possibility is that the model city award does not generate any significant air quality improvement throughout the year but it generates incentives to game around the blue sky threshold when a city faces higher pressure to reach the target toward the end of a year. If we rank cities according to their average API in year 2001 only and then correlate this rank with the order of receiving the model city award, we find the spearman rank correlation to be 0.42 (significant at 99% confidence). This leads us to wonder whether the model city award was designed simply to recognize cross-sectional variation of air quality across cities rather than to encourage further environmental protection efforts within a city.

It is also possible that local cities adopt real measures to *temporarily* improve air quality around the threshold of blue sky days at the time of winning the model city award. However, the reported API is supposed to be an average across multiple stations and multiples times for a major pollutant on a city-day. It

is difficult to aim for a particular integer after such an arithmetic aggregation. Also, the translation from city-wide anti-pollution measures to pollutant density is noisy, not immediate, and difficult to predict beforehand. Some temporary improvements — for example, relocating a monitoring instrument to a cleaner area — are gaming, by our definition.

The third possibility is that gaming around the blue sky threshold is dominated by overall air quality changes before, during, and after a city wins the model city award. If every city improves in response to the model city policy regardless of its award-winning status, such improvement contributes to national variation. This could explain why we do not find any statistical improvement after controlling for date fixed effects nationwide. But this explanation is questionable if we consider how lax re-examination is after a city wins the model city award. Other facts to be considered are that air quality is only one of the many statistics used in model city evaluation and that model city evaluation is only one of the many tools that the central government may use to promote local government leaders. These considerations imply that cities could have improved air quality in response to other policies that provide stronger and more continuous incentives than the model city award. Identifying such policies is a potential direction for future research.

## 9. References

- Andrews, Steven (2008) “Inconsistencies in air quality metrics: ‘Blue Sky’ days and  $PM_{10}$  concentrations in Beijing” *Environmental Research Letters*, 2008(3): 034009.
- Becker, Randy and Vernon Henderson (2000) “Effects of Air Quality Regulations on Polluting Industries” *The Journal of Political Economy*, 108 (2): 379-421.
- Brook, Dann and Burnett (1997) “The relationship among TSP,  $PM_{10}$ ,  $PM_{2.5}$ , and inorganic constituents of atmospheric particulate matter at multiple Canadian locations” *Journal of the Air & Waste Management Association* ISSN1096-22471997, vol. 47, n<sup>o</sup>1, pp. 2-19 (44 ref.)
- Burgstahler, D. and I. Dichev (1997) “Earnings management to avoid earnings decreases and losses” *Journal of Accounting and Economics*, 24, 99-126.
- Chay, Kenneth and Michael Greenstone (2003) “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession” *Quarterly Journal of Economics*, 118(3).
- Chay, Kenneth; Carlos Dobkin and Michael Greenstone (2003) “The Clean Air Act of 1970 and Adult Mortality” *Journal of Risk and Uncertainty*, 27(3).
- Che, Huizheng; Xiaoye Zhang; Yang Li; Zijiang Zhou; and Zhenline Chen (2006) “Relationship Between Horizontal Extinction Coefficient and  $PM_{10}$

- Concentration in Xi'An, China During 1980-2002" *China Particuology* 4(6): 327-329.
- Chen, Ye; Hongbin Li and Li-An Zhou (2005) "Relative Performance Evaluation and the turnover of Provincial Leaders in China" *Economics Letters*, 88(3): 421-25.
- Chen, Yuyu; Ginger Zhe Jin; Naresh Kumar and Guang Shi (2011) "The Promise of Beijing: Evaluating the Impact of the 2008 Olympic Games on Air Quality", NBER working paper #16907.
- Congleton, R. (1992) "Political Institutions and Pollution Control" *Review of Economics and Statistics* 74: 412-21.
- Chu, D. A.; Y. J. Kaufman; G. Zibordi; J. D. Chern; J. Mao; C. Li; and B. N. Holben (2003) "Global Monitoring of Air Pollution over Land from EOS-Terra MODIS." *Journal of Geophysical Research*, 2003. 108(D21): p. 4661.
- Currie, Janet, and Matthew Neidell (2005) "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" *Quarterly Journal of Economics* 120(3): 1003-30.
- Currie, Janet; Matthew Neidell and Johannes F. Schmieder (2009) "Air pollution and infant health: Lessons from New Jersey" *Journal of Health Economics*, 28(3): 688-703.
- Davis, Lucas W. (2008) "The Effect of Driving Restrictions on Air Quality in Mexico City" *Journal of Political Economy* 116(1): 37-80.
- Dean, J.M.; Lovely, M.E. and Wang, H. (2009) "Are foreign investors attracted to weak environmental regulations? Evaluating the evidence from China" *Journal of Development Economics*, 90(1): 1-13.
- DeGeorge, Francois; Jayendu Patel; and Richard Zeckhauser (1999) "Earnings Management to Exceed Threshold." *Journal of Business*, 72(1): 1-33.
- Ederington, J.; Levinson, A. and Minier, J. (2005) "Footloose and Pollution-Free" *The Review of Economics and Statistics* 87(1), 92-99.
- Edin, Maria (2003) "State Capacity and Local Agent Control in China: CCP Cadre Management from a Township Perspective" *China Quarterly*, 173: 35-52.
- Fan, Yin-Qi and Chun Li (2008) "Visibility Trends in Beijing, Tianjin and Hebei Province During 1980-2003" *Plateau Meteorology* 27(6): 1392-1400.
- Fisman, Raymond, and Shang-Jin Wei (2004) "Tax Rates and Tax Evasion." *Journal of Political Economy*, 112: 471-496.
- Fisman, Raymond and Shang-Jin Wei (2007) "The Smuggling of Art and the Art of Smuggling: Uncovering the Illicit Trade in Cultural Property and Antiques," NBER Working Paper No. W13446.
- Forbes, Silke; Mara Lederman and Trevor Tombe (2011) "Quality Disclosure Programs with Thresholds: Misreporting, Gaming and Employee

- Incentives” *working paper*, accessed at <http://www.econ.ucsd.edu/~sjanusze/>.
- Greenstone, Michael (2002) “The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures” *Journal of Political Economy*, 110(6): 1175-1219.
- Greenstone, Michael (2004) “Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations?” *Journal of Environmental Economics and Management*, May 2004.
- Guinot, Benjamin (2008) “Olympic Games: Beijing Air Quality” accessed at [http://www.ambafrance-cn.org/奥林匹克运动会：北京大气质量\\_4570.html](http://www.ambafrance-cn.org/奥林匹克运动会：北京大气质量_4570.html) on January 10, 2010.
- Gupta, Pawan; Sundar A. Christopher; Jun Wang; Robert Gehrig; Yc Lee; Naresh Kumar (2006). "Satellite remote sensing of particulate matter and air quality assessment over global cities." *Atmospheric Environment* 40(30): 5880-5892.
- Henderson, J. Vernon (1996) “Effects of Air Quality Regulation” *The American Economic Review*, 86 (4): 789-813.
- Huang, Wei; Jianguo Tan; Haidong Kan; Ni Zhao; Weimin Song; Guixiang Song; Guohai Chen; Lili Jiang; Cheng Jiang; Renjie Chen; and Bingheng Chen (2009) “Visibility, air quality and daily mortality in Shanghai, China” *Science of the Total Environment*, 407 (2009): 3295–3300.
- Jacob, Brian and Steven Levitt (2003) "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, 118(3): 843-877.
- Jiang, D.; Y. Zhang; X. Hu; Y. Zeng; J. Tan; and D. Shao (2004) “Progress in developing an ANN model for API forecasting” *Atmospheric Environment*, 28: 7055-64.
- Kaufman, Yoram J.; Nadine Gobron; Bernard Pinty; Jean-Luc Widlowski; and Michel M. Verstraete (2002). "Relationship between surface reflectance in the visible and mid-IR used in MODIS aerosol algorithm - theory." *Geophysical Research Letters* (0094-8276)29(23): 31-34.
- Kaufman, Yoram J., D. Tanre; B.N. Halben; S. Mattoo; L.A. Remer; T.F. Eck; J. Vaughan and B. Chatenet (2002). "Aerosol radiative impact on spectral solar flux at the surface, derived from principal-plane sky measurements." *Journal of the Atmospheric Sciences* 59(3): 635-646.
- Keller, W. and Levinson, A. (2002) “Pollution Abatement Costs and Foreign Direct Investment Inflows to U.S. States” *The Review of Economics and Statistics*, 84(4): 691-703.

- Kumar, Naresh and Andrew Foster (2009), "Air quality interventions and spatial dynamics of air pollution in Delhi and its surroundings," *International Journal of Environment and Waste Management*, 4(1/2): 85-111.
- Kumar, N., A. Chu, and A. Foster (2008) "Remote sensing of ambient particles in Delhi and its environs: estimation and validation." *International Journal of Remote Sensing*, 29(12): p. 3383-3405.
- Kumar, N., A. Chu, and A. Foster (2007) "An Empirical Relationship between PM<sub>2.5</sub> and Aerosol Optical Depth in Delhi Metropolitan." *Atmospheric Environment*, 41(21): p. 4492-4503.
- Kumar, N. (2010). "A Hybrid Approach for Predicting PM<sub>2.5</sub> Exposure." *Environmental Health Perspective*, 118(10).
- Kumar, N., A. Foster, S.N. Tripathi, A. Dasguptai, and A. D. Chu (2009) "Predicting Air Quality using MODIS data", [web.ccs.miami.edu/~nkumar/EST\\_Final\\_Version.pdf](http://web.ccs.miami.edu/~nkumar/EST_Final_Version.pdf).
- Kumar, Naresh; Allen D. Chu; Andrew D. Foster; Thomas Peters; and Robert Willis (2011). "Satellite Remote Sensing for Developing Time and Space Resolved Estimates of Ambient Particulate in Cleveland, OH." *Aerosol Science and Technology* 45(9): 1090-1108.
- Levitt, Steven D. and Chad Syverson (2008) "Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions," *Review of Economics and Statistics* 90(4): 599-611.
- Levy, R. C., L. A. Remer, et al. (2007a). "Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land." *Journal of Geophysical Research-Atmospheres* 112(D13).
- Levy, R. C., L. A. Remer, et al. (2007b). "Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance." *Journal of Geophysical Research-Atmospheres*, 112(D13).
- List, John A.; Daniel L. Millimet; Per G. Fredriksson and W. Warren McHone (2003): "Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator" *The Review of Economics and Statistics*, 85(4): 944-952.
- Maskin, Eric; Yingyi Qian and Chenggang Xu (2000): "Incentives, Information, and Organizational Form." *Review of Economic Studies*, 67(2): 359-78.
- McCarry, Justin (2008) "Manipulation of the Running variable in the Regression Discontinuity Design: A Density Test" *Journal of Econometrics*, 142(2), February.
- Michaely, Roni, and Kent Womack (1999) "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations," *Review of Financial Studies*, 12 (4): 653-686.



- Mishra, Prachi, Petia Topalova, and Arvind Subramanian (2007) "Policies, Enforcement, and Customs Evasion: Evidence from India," IMF Working Paper No 07/60.
- Murdoch, J. and T. Sandler (1997) "The Voluntary Provision of a Pure Public Good: The Case of Reduced CFC Emissions and the Montreal Protocol" *Journal of Public Economics* 63: 331-50.
- NASA. (2010). "the Level 1 and Atmosphere Archive and Distribution System." Retrieved July 3, 2010, from <http://ladsweb.nascom.nasa.gov/>.
- NCDC. (2007). "National Climatic Data Center." from <http://www.ncdc.noaa.gov/oa/ncdc.html>.
- Oats, Wallace and Paul Portney (2003) "The Political Economy of Environmental Policy" in K.G. Maler and J. Vincent edited Handbook of Environmental Economics (Amsterdam: Elsevier) pp. 325-354.
- Ozkaynak H, et al. (1985) "Relationship between aerosol extinction coefficients derived from airport visual range observations and alternative measures of airborne particle mass" JAPCA. 35 (1985):1176-85.
- Saez, Emmanuel (2010) "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2: 180-212.
- Simonich, Stacil L. Massey (2009) "Response to Comments on 'Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics'" *Environmental Science & Technology* (43): 7590-7591.
- Slemrod, Joel (1985) "An Empirical Test for Tax Evasion," *Review of Economics and Statistics* 67(2): 232-238.
- Snyder, Christopher and Owen Zidar (2009). "Resume Padding Among Economists?" Unpublished.
- Takeuchi, Y (2004) "On a statistical method to detect discontinuity in the distribution function of reported earnings" *Mathematics and Computers in Simulation*, 64(1): 103-111.
- Tang, Xiaoyan; Min Shao; Min Hu; Zifa Wang; and Junfeng (Jim) Zhang (2009) "Comment on 'Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics'" *Environmental Science & Technology* (43): 7588.
- Tsui, Kai-yuen and Youqiang Wang (2004) "Between Separate Stoves and a Single Menu: Fiscal Decentralization in China" *China Quarterly*, 177: 771-90.
- United Nations Environment Programme (2009) "Independent Environmental Assessment Beijing 2008 Olympic Games." Accessed at [http://www.unep.org/publications/UNEP-eBooks/BeijingReport\\_ebook.pdf](http://www.unep.org/publications/UNEP-eBooks/BeijingReport_ebook.pdf).
- Viard, V. Brian and Shihe Fu (2011): "The Effect of Beijing's Driving Restrictions on Pollution and Economic Activity" *working paper*, accessed at <http://www.ckgsb.edu.cn/personalsites/brianviard/papers.htm>.

- Wang, Wentao; Toby Primbs; Shu Tao; and Stacil L. Massey Simonich (2009) "Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics" *Environmental Science & Technology* (43): 5314-5320.
- Wheeler, D. (2004) "Racing to the Bottom? Foreign Investment and Air Pollution in Developing Countries" *The Journal of Environment and Development* 10(3): 225-245.
- Whiting, Susan H. (2000) *Power and Wealth in Rural China: The Political Economy of Institutional Change*. Cambridge University Press.
- Xu, Chenggang (2011) "The Fundamental Institutions of China's Reforms and Development" *Journal of Economic Literature*, 49(4): 1076-1151.
- Yao, Xiaohong; Xiaohong Xu; Kelly Sabaliauskas and Ming Fang (2009) "Comment on 'Atmospheric Particulate Matter Pollution during the 2008 Beijing Olympics'" *Environmental Science & Technology* (43): 7589.
- Zeng, D. Z. and Zhao, L. (2009) "Pollution havens and industrial agglomeration" *Journal of Environmental Economics and Management* 58(2): 141-153.
- Zinman, Jonathan and Eric Zitzewitz (2009) "Wintertime for Deceptive Advertising?" <http://www.dartmouth.edu/~ericz/wintertime.pdf>.
- Zitzewitz, Eric (2012) "Forensic Economics" *Journal of Economic Literature* 50(3): 731-69.