'Effortless Perfection:' Do Chinese Cities Manipulate Air Pollution Data?*

Dalia Ghanem† Junjie Zhang ‡

University of California, San Diego

April 25, 2013

Abstract

It is alleged that some Chinese cities manipulate their air pollution data to comply with the air quality standards set by the central government of China. This paper tests this hypothesis by using a unique data set. First, we employ a discontinuity test to detect the cities that reported dubious pollution data around the cutoff for "blue-sky days." Then, we propose a panel matching approach to identify the conditions under which irregularities may occur. Over the period 2001-2010, 55% of cities reported dubious PM_{10} pollution levels that led to a discontinuity at the cutoff. Suspicious data reporting tends to occur on days when the potential manipulation is least detectable. Our findings indicate that the official daily air pollution data are not well behaved, which provides suggestive evidence of manipulation.

Keywords: air pollution, manipulation, discontinuity test, panel matching.

JEL Classification Numbers: Q53, C23, C14.

*The authors thank the UC Center for Energy and Environmental Economics (UCE^3) for partial financial support. We thank Roger Bohn, Richard Carson, Graham Elliott, Josh Graff-Zivin, Mark Jacobsen, David Meyer, Steven Oliver, and Logan Smith for their helpful suggestions. We also benefited from the comments and suggestions from the seminar discussants and participants at UCSD and AERE meetings. Of course, all remaining errors are ours.

†Department of Economics, University of California, San Diego. dghanem@ucsd.edu.

‡Corresponding Author. School of International Relations & Pacific Studies, University of California, San Diego. 9500 Gilman Drive #0519, La Jolla, CA 92093-0519. Tel: (858) 822-5733. Fax: (858) 534-3939. junjiezhang@ucsd.edu.

1 Introduction

Air quality of major Chinese cities is among the worst in the world, a consequence of three decades of double-digit economic growth with lax environmental regulation (Zhang and Crooks, 2012). The central government of China started to address the issue of environmental regulation. Over a decade ago, major Chinese cities were required to report their daily air pollution index (API) and to make it available to the public. To incentivize air pollution abatement, performance evaluations of local officials included the number of "blue-sky days," which are days with API below 100.[1] These measures introduced both bottom-up and top-down pressures on the local governments to meet the target number of blue-sky days. In the absence of independent verification mechanisms, some cities are allegedly under-reporting their air pollution levels. The presence of a particular threshold for blue-sky days seems to be associated with dubious behavior of the API scores around the cut-off, which is illustrated in Figure 1. We call this phenomenon "effortless perfection." Specifically, the API distributions of three municipalities, including Beijing, Tianjin and Chongqing, amass right below the cut-off. The irregularities have raised suspicion of systematic manipulation of air pollution data in China.

At the heart of this paper is a question about self-reported data that are used in performance evaluations of the reporting agencies. In such settings, there is good reason to suspect that the agencies manipulate the data to improve their performance evaluations, especially if such manipulation can go undetected. The relevance of this paper is hence not restricted to air pollution reporting. Using the unique data set that we have, our main goal is to understand if and how manipulation occurs in this setup. The importance of such a question goes beyond detecting cheaters. It can help us understand how manipulation can undermine the use of such self-reported data in economic studies.

Data manipulation is a concern for both policy makers and researchers. For policy makers, data manipulation defeats the purpose of policies that incentivize pollution abatement and other policy objectives. It might be effortless for a city to beautify its environmental record by just tweaking the numbers. However, society must then bear the costly consequence of misleading pollution data. First, misinformed urban residents cannot efficiently mitigate pollution-related health effects

[1] A blue-sky day is associated with low or moderate health concerns. See Table 3 for categories of API and health concerns.

by changing their behavior, for example wearing masks or canceling outdoor activities. Second, promotion of officials whose cities have become involved in manipulation jeopardizes the public interest because their integrity is in question. In addition, environmental compliance through manipulation undermines the government's credibility.

Manipulation introduces measurement error into the data that may be used as regressors to study the effect of pollution on health and other outcomes. Such measurement error will bias studies that evaluate the impact of air pollution using such data on various outcomes.[2] The biased marginal effects of pollution might lead to incorrect policy recommendations. If the measurement error in the air pollution data violates the classical measurement-error assumptions, then standard econometric methods may not rectify the bias due to the measurement error in this data.

Our paper aims to identify irregularities in the air pollution data in order to provide insight on the nature of manipulation and the circumstances under which it occurs. We define manipulation as the behavior of not reporting the true pollution level, such as data falsification or hiding bad pollution data. It does not include strategic behavior such as temporary driving bans, closing factories, or requiring different fuels.[3] Although these command-and-control policies are inefficient, they can indeed reduce pollution in the short run so we cannot call them manipulation. The best way to detect manipulation is to use alternative sources of data to validate the official data. However, independent direct measures are not available for all the relevant pollutants, cities, and periods.[4] Therefore, this paper uses econometric methods to uncover suggestive evidence of manipulation from the self-reported data.

We pose two research questions. First, we investigate whether a city reports dubious pollution data around the cutoff for blue-sky days.[5] To answer this question, our empirical strategy is borrowed from the McCrary (2008) test, which is used to detect manipulation of a running variable

[2] A large body of literature in environmental and health economics studies the impact of air pollution on productivity (Koop and Tole, 2004), property values (Chay and Greenstone, 2005), and health (Graff Zivin and Neidell, 2012). The reliability of the estimated marginal effects hinges on valid pollution data.

[3] These policies were used during major events such as 2008 Olympic Games in Beijing and Expo 2010 in Shanghai. They are also used when local governments are desperate to meet their environmental targets at the end of a particular year.

[4] A notable independent measure is the U.S Embassy Beijing Air Quality Monitor. It has reported $PM_{2.5}$ particulates pollution since 2008. However, $PM_{2.5}$ was not regulated during 2001-2010. Some research institutions collected air quality data independently but the data are not widely available.

[5] Moral hazard arises where there is a cut-off in the performance evaluation and the information is asymmetric. Manipulation around the cut-off is not a unique phenomenon in economics. For example, lenders treat mortgage borrowers with credit scores just above certain thresholds differently from those with slightly lower scores. Borrowers may take advantage of this rule of screening by manipulating their FICO scores (Keys et al., 2010).

in regression discontinuity design. The intuition here is that if the pollutant concentration on a particular day misses the cutoff for blue-sky day by a small amount, then there is an incentive for the city to under-report the pollutant concentration and score a blue-sky day. If such behavior occurs often, the distribution of the pollutant concentration exhibits a discontinuity at the cutoff for a blue-sky day. In the absence of manipulation, the distribution of air pollutant concentrations is expected to be continuous because polluters and regulators do not have complete control over the realized pollutant concentration (Brannlund and Lofgren, 1996). Thus, detection of this type of manipulation boils down to a test of discontinuity around the cutoff for blue-sky days in the distribution of pollutant concentrations. Irregularity around the cut-off is a red flag of potential manipulation.[6]

Second, we study the patterns of manipulation by proposing a panel matching approach. The ideal experiment to examine such patterns is to observe twin cities that are expected to have identical distributions of air quality, and hence of the API. The panel matching approach constructs pairs of cities that have the same geographic and provincial characteristics. Since true air quality is unobservable, we use visibility as a proxy together with other weather variables (Sloane and White, 1986). The key assumption that allows us to identify manipulation is that for a constructed city pair, we expect that the two cities have the same distribution of API conditional on visibility and other weather variables. This approach does not pin down the manipulator.[7] Instead, it identifies the conditions under which manipulation is most likely to occur.

This paper is not the first attempt to investigate dubious air pollution data in China. Andrews (2008a,b) first questioned the credibility of official data in Beijing and brought this issue to public attention. The author presented the evidence that the API has massive bunching below the cut-off in addition to other gimmicks to polish air quality reports. Chen et al. (2012) provide a formal econometric analysis on the accuracy of the air pollution data. They confirmed the anomaly around the cut-off based on the official data from 37 large cities during 2000-2009. We improve on this literature in three aspects.

First, we use a more comprehensive data set. In particular, we obtained the confidential daily

[6] As noted above, command-and-control policies can also be used to ensure that the target number of blue-sky days is met. Such behavior can also lead to a bunching below the cutoff for blue-sky days.

[7] Inspection of the empirical CDFs can suggest which city is the manipulator, but the formal tests we perform do not answer this question.

air pollution data from the China National Environmental Monitoring Center (CNEMC). The data include non-disclosed variables underlying the calculation of the API. Our data set covers 113 cities during 2001-2010, which includes all cities that are required to disclose daily air pollution information. The detailed data have never been used in previous studies.

Second, our data set allows us to apply the discontinuity test to pollutant concentrations directly instead of the API. The distribution of pollutant concentrations satisfies the continuity assumption of the McCrary (2008) test, hereinafter the McCrary test, whereas that of the API does not. The API is a nonlinear transformation of pollutant concentrations. Its distribution is not continuous, which violates the assumption required for the McCrary test. Applying the McCrary test to the API scores directly may lead to biased results.

Third, our panel matching method is novel. The nonparametric specification is superior to a linear model because the former does not assume that manipulation and control variables such as visibility and weather conditions are separable, which allows for more realistic forms of manipulation which differ based on weather conditions. Furthermore, linear fixed effects models are known to be inconsistent when they are misspecified as shown in Chernozhukov et al. (2010) and Gibbons, Serrato, and Urbancic (2011). Since we do not expect the true relationship between API and other variables to be linear or to follow a particular functional form assumption, the nonparametric approach is preferable.

Our primary finding is that 55% of cities report dubious particulate matter (PM_{10}) pollution data. In comparison, the data for sulfur dioxide (SO_2) and nitrogen dioxide (NO_2) are much more trustworthy. This is not surprising, since PM_{10} is the primary culprit of non-blue-sky days in the majority of Chinese cities. In terms of the circumstances under which manipulation occurs, we find that dubious data tend to occur on days with high visibility and low wind speed. When wind speed is low, nature is not doing its part and the pollutants are not simply "gone with the wind." On days with high visibility, manipulation is not easily detectable. These findings are very intuitive and illustrate the potential manipulation behavior of local governments. It is important to note that the two methods we use to detect manipulation may be applied by the monitoring agencies themselves to detect manipulators. Although we focus on air pollution, the methodology extends to other situations where the reported data are subject to manipulation due to the presence of a particular cutoff.

The rest of the paper is organized as follows. Section 2 gives an overview of the empirical setting. Section 3 describes the data and variables. Section 4 uses a discontinuity test to detect irregularities around the cut-off. Section 5 proposes the panel matching approach to identify the pattern of manipulation. Section 6 presents robustness checks and further discussions, and Section 7 concludes. Additional estimates and robustness checks are included in the online Appendix.

2 Empirical Background

This section introduces air pollution and its regulation in China. We focus on the institutional background on why some Chinese cities may engage in data manipulation.

2.1 Air Pollution and Regulation

Air pollution is a critical environmental concern in China. Asia Development Bank reports that less than 1% of the largest Chinese cities meet the air quality standards recommended by the World Health Organization (Zhang and Crooks, 2012). With significant efforts to curb pollution, some pollutant emissions have slowed down and have not reached their peaks in recent years. For example, total sulfur dioxide emissions peaked in 2006 at 25.89 million metric tons, when the per-capita GDP was $2,064 USD in current prices.[8] Nonetheless, China is still the world's largest emitter of many air pollutants including sulfur dioxide and nitrogen oxide. Poor air quality is a result of rapid economic growth that heavily relies on fossil fuel consumption (Zhang and Wang, 2011). Coal accounts for about 70% of total energy use, which has led to severe SO_2, NO_2, and particulate matter pollution. In addition, motor vehicle usage has grown dramatically, since private car ownership has increased from 3.43 million in 2002 to 78.72 million in 2011.[9] Automotive consumption of gasoline has become a major source of air pollution in big cities.

Severe air pollution has caused tremendous health, economic, environmental, and social problems. Because of SO_2 and NO_2 emissions, acid rain occurred in 227 cities in 2011, or about half of all the monitored cities.[10] Although particulate-matter pollution has improved significantly

[8]National Bureau of Statistics, 2012. China Statistical Yearbook. http://www.stats.gov.cn/tjsj/ndsj/ (Retrieved on Oct 7, 2012).

[9]National Bureau of Statistics, 2003-2011. Statistical Communique on the National Economic and Social Development. http://www.stats.gov.cn/tjgb/ (Retrieved on Oct 7, 2012).

[10]Ministry of Environmental Protection. 2012. "Report on the State of the Environment in China in 2011." http:

since 2005, its concentrations are still five times higher than the safety level. A World Health Organization (WHO) report estimates that each year 656,000 premature deaths of Chinese citizens are attributed to the diseases triggered by air pollution.[11] Another recent study suggests that the welfare loss caused by ozone and particulate-matter pollution in 2005 is about 112 billion of 1997 U.S. dollars (Matus et al., 2012). The monetized health costs of air pollution alone are estimated to be between 1.2% and 3.8% of GDP (World Bank, 2007).[12] In addition, pollution has stirred widespread discontent among the emerging middle class in urban areas, resulting in what the Chinese government defines as "mass incidents." These mass incidents have threatened social stability that is regarded as a top priority for the Chinese central government.[13] They have also created bottom-up pressures for local governments to clean up the environment.

In the wake of serious air pollution, China has been constructing a national system of atmospheric air pollution standards since 1982 (Natural Resources Defense Council, 2009). A history of the standards and the criteria air pollutants is summarized in Table 1. As a national strategy to improve ambient air quality, 113 key cities are required to disclose their once classified air quality data. The mandate began with weekly reporting in 1998 and advanced to daily reporting in 2000. The central government uses information disclosure to create an incentive for local governments to engage in air pollution reduction more actively. Disclosed air pollution data are used not only to inform the public but also to evaluate city officials' environmental performance.

However, China's air pollution regulations have faced a fair amount of critique: the regulations are relatively lax compared to the standards recommended by WHO or those adopted by other developed countries; certain pollutants are not included; and in some cases the standards have been revised downward to increase compliance. The most pronounced case took place in the 2000 revision. In response to non-compliance due to the increase in automobile usage, the regulator removed NO_x from the list of the criteria pollutants. The standards for NO_2 and ozone (O_3) were lowered as well. Due to the lax standard, ambient NO_2 concentrations are seldom considered

//jcs.mep.gov.cn/hjzl/zkgb/.

[11] Kevin Holden Platt. July 9, 2007. Chinese Air Pollution Deadliest in World, Report Says. http://news.nationalgeographic.com/news/2007/07/070709-china-pollution.html.

[12] The World Bank used the adjusted human capital (AHC) approach to estimate the forgone earnings due to pollution at 1.2% of GDP. They used the value of a statistical life (VSL) approach to estimate the mortality risks at 3.8% of GDP.

[13] Junjie Zhang. November 9th, 2012. How Will China's New Leaders Approach Rising Tide of Environmental Protests? http://asiasociety.org/blog/asia/how-will-chinas-new-leaders-approach-rising-tide-environmental-protests.

the primary air pollutant. Another important case is $PM_{2.5}$, a fine particulate with major health consequences, which was not included in the standards until 2012. Fortunately, during the period of our study, 2001-2010, the standards remained consistent.

2.2 Costly Action vs. Effortless Perfection

Although China has a relatively comprehensive system of air pollution regulation, implementation of the standards at the local level is a major problem. In order to motivate local officials to reduce pollution, environmental compliance has entered the cadre promotion system.[14] Specifically, 113 key cities have been ranked in the annual Quantified Assessment of Urban Environmental Improvement (*Chenkao*) since 1989. Air quality is the single most important indicator in the assessment, which accounts for 20 percent of a city's environmental quality grade. During the 11th Five-Year Plan period (2006-2010), a city with automated monitoring systems receives 20 points if the annual count of blue-sky days is greater than 85% of a year and 0 points if the share is less than 30%. In other cases, the city's grade is determined by $20 \times (p - 30\%)/55\%$, where p is the proportion of blue-sky days.[15] Therefore, city officials strive to achieve 85% of blue-sky days in a year in order to obtain a full score on air quality.

Local officials are expected to comply with the environmental standards because their prospects for career advancement are linked to their ability to meet the targets set by the higher-level offices. In addition, local officials compete with each other on observable performance measures including economic output and social stability, creating a promotion tournament (Chen, Li, and Zhou, 2005; Li and Zhou, 2005; Shih, Adolph, and Liu, 2012). The rankings of environmental performance by the *Chenkao* were intended to award the title of "Environmental Protection Model City" to top performing cities. The yardstick competition that it creates among city mayors was intended to improve the environment. Zheng et al. (2013) provide empirical evidence that Chinese mayors' likelihood of promotion is affected by both economic growth rate and environmental performance among other things.

However, some performance indicators are difficult to monitor and verify. The central gov-

[14] Ian Johnson. June 3, 2011. China Faces "Very Grave" Environmental Situation, Officials Say. http://www.nytimes.com/2011/06/04/world/asia/04china.html (Retrieved on November 29, 2012).

[15] More details about the *Chenkao* (in Chinese) are available at: http://wfs.mep.gov.cn/chengkao/ckzb/200612/P020061229356985550756.pdf (Retrieved on February 6, 2013).

ernment must often rely on data that are self-reported by local governments. Under asymmetric information, local officials have an incentive to use inappropriate behavior if their interests are not aligned with those who grant promotions. In a worst-case scenario, those officials may engage in data manipulation. Credibility of official statistical data in China has already been under international scrutiny, most notably the overstated economic growth rate (Rawski, 2001; Koech and Wang, 2012). This has led to similar concerns about the integrity of the environmental data.

Although environmental compliance has been explicitly written into the contract between the central and local governments, economic growth is still regarded as the top priority in China (Zhang, 2012). Local officials have unparalleled enthusiasm for growing the economy because of the dual incentives of financial rewards and political futures. This is consistent with the argument that authoritarian leaders opt for less environmental goods in return for faster economic growth (Congleton, 1992). Local governments might lower environmental standards in order to appeal to investors and raise competitiveness, creating a "race to the bottom." Even worse, taking advantage of the asymmetric information between the central and local governments, self-interested local officials might overstate economic achievement and understate environmental pollution.

3 Data Description

We have assembled a unique data set for the empirical analysis, which integrates a confidential air pollution data set with visibility and other weather conditions.

3.1 Daily Air Pollution

The air pollution data are provided by the China National Environmental Monitoring Center (CNEMC), which is affiliated with the Ministry of Environmental Protection of China. Note that CNEMC faithfully compiled the air pollution data reported by the local governments during 2001-2010. Since CNEMC is neutral with respect to local interests, we are relatively certain that any anomalies in the data are attributable to the local governments.[16]

The air pollution data contain two parts. The first part is public, which includes daily API score and primary pollutant. The main shortfall of the public data is that pollutant concentra-

[16]The judgment is based on personal communication with officials from CNEMC.

tions are not reported. Although the primary pollutant concentration can be inferred from its API score, non-primary pollutants' information is unknown. In addition, the primary pollutant is not reported in a non-pollution day (API ≤ 100). Fortunately, we have obtained the confidential part that includes concentrations of all three criteria pollutants: PM_{10}, SO_2, and NO_2. The confidential pollutant concentration information has never been used in previous studies.

The air quality data cover 113 cities from 2001 to 2010. The spatial distributions of air pollution in terms of API and three criteria pollutant concentrations are illustrated in Figure 2. The spatially interpolated air pollution levels shown by the filled contour plots are generated by inverse distance weighting approach based on the city-level daily air pollution data. Air pollution, particularly PM_{10} pollution, is generally worse in North and Northwest China. It is caused by a combination of pollution, geographic and meteorological conditions.

The summary statistics are reported in Table 4. PM_{10} was the dominant primary pollutant in all cities, responsible for 73.7% of non-blue-sky days (API ≥ 100). SO_2 caused less than 10% of non-blue-sky days. NO_2 was almost never responsible for non-attainment because of its lax standard. The mean API is 76.32, implying the average air quality meets the requirement of blue-sky days. On average, blue-sky days account for 84.6% days in the last decade, which interestingly coincides with the target set by the central government. Days with Grade II air quality accounted for 67.7%, that is, most days had good air quality with moderate health consequences.

3.2 Air Pollution Index (API)

Air quality is reported in the form of both pollutant concentrations and Air Pollution Index (API). API converts concentrations of three criteria air pollutants (PM_{10}, SO_2, and NO_2) to a single index by a set of piece-wise linear transformations. The index I for one pollutant with concentration C is defined as the linear interpolation between two index classes such that

$$I = \frac{I_u - I_l}{C_u - C_l}(C - C_l) + I_l. \tag{1}$$

In this form, C_u and C_l are the upper and lower boundaries of concentrations for each air quality level, and I_u and I_l are the corresponding upper and lower index classes. The thresholds are reported in Table 2. Although nine air pollutants were regulated during this period, only three

pollutants enter the daily air pollution report system because of technical and cost constraints. The normalized index for each pollutant is computed based on its daily average concentration.

API on a given day is determined by the pollutant that has the highest index. The corresponding air pollutant is referred to as the primary pollutant.

$$API = \max\{I_{\mathrm{SO}_2}, I_{\mathrm{NO}_2}, I_{\mathrm{PM}_{10}}\}. \tag{2}$$

API varies between 0 and 500 with a large number indicating poor air quality. Different API categories are associated with different pollution levels and health consequences (Table 3). Officially, a "blue-sky day" is defined as a day with the value of API less than 100, that is, the air quality is either excellent or good.[17] The compliance with the air quality standards is then simplified by just counting the number of blue-sky days.

Pollutant concentrations are measured and averaged across stations and over a 24-hour period. In order to release pollution information in the afternoon, daily report uses the data from the previous noon to current noon. To summarize, API calculation is implemented in four steps: First, a 24-hour average pollutant concentration is calculated for each station. Secondly, city average pollutant concentration is derived from multiple station averages. Thirdly, individual pollutant index I is calculated according to equation (1). And finally, API is the maximum of individual pollutant indices according to equation (2).

Pollutant concentrations are measured and averaged across stations and over a 24-hour period. In order to release pollution information in the afternoon, daily report uses the data from the previous noon to current noon. To summarize, API calculation is implemented in four steps: First, a 24-hour average pollutant concentration is calculated for each station. Secondly, city average pollutant concentration is derived from multiple station averages. Thirdly, individual pollutant index I is calculated according to equation (1). And finally, API is the max of individual pollutant index according to equation (2).

Manipulation happens in the process of calculating daily average pollutant concentrations at station- or city-level. It could be caused by data falsification, which is against the law. It could also be caused by regulatory loopholes. Specifically, the minimum data requirement to calculate

[17]Note that a blue-sky day is a just technical definition. It does not necessarily mean that the sky is literally blue.

daily averages for gaseous pollutants (SO_2 and NO_2) is 18 hours of effective monitoring and that for particulate matter is 12 hours.[18] This creates an incentive for cities to discard the observations with bad pollution on the excuse of faulty equipment. In addition, since the data requirement for PM_{10} is lower than the other two pollutants, this becomes another reason why PM_{10} is more vulnerable to manipulation besides PM_{10} being the dominant primary pollutant.

3.3 Visibility and Other Weather Variables

The meteorological data are obtained from the National Climatic Data Center under the National Oceanic and Atmospheric Administration (NOAA) of the United States. The weather data are collected by the weather stations under the China Meteorological Administration (CMA). Since weather stations are not prone to political interference, the human operators of weather stations do not have an incentive to manipulate the results. This data set that we use records weather information for 499 weather stations every three hours from 2001-2010. Note that we dropped the weather stations that have less than 10,000 records.

Meteorological factors are correlated with air pollution. The variables that we use in the paper include visibility (VSB, in statute miles), temperature (TEMP, in Fahrenheit), atmospheric pressure (STP, in millibars), precipitation (PCP, in inches), and wind speed (SPD, in miles per hour). The weather variable of central interest is visibility, which is used as a proxy for API. Visibility is historically defined as "the greatest distance at which an observer can just see a black object viewed against the horizon sky" (Malm, 1999). It has been shown that particulate matter and gaseous pollution can cause visibility impairment. All these variables are daily averages in order to match the time scale of API data. The summary statistics for weather variables are also reported in Table 4. In particular, average visibility is 6.5 miles.

4 Irregularities at the Cutoff for Blue-Sky Days

One expects that manipulation around the cutoff for blue-sky days is the most likely form of manipulation to occur. The reasoning here is straightforward. Local governments report their

[18] Automated Methods for Ambient Air Quality Monitoring HJ/T193-2005. http://www.zhb.gov.cn/image20010518/5523.pdf

API scores on a daily basis and this data set is publicly available. If the API scores are tweaked by large amounts, citizens and central government officials may doubt the information reported by local governments. Manipulation right around the cut-off is less likely to be detected because the difference in air quality between API values at 100^- and 100^+ may be indiscernible. Hence, we may rightly predict that cities manipulate blue-sky days by examining the discontinuity in the probability density function (pdf) of air quality data around the cutoff.

4.1 API vs Concentration

McCrary (2008) proposes a test for the manipulation of the running variable in regression discontinuity design. An implication of the manipulation of the running variable is that its pdf would have a discontinuity. The main assumption for the validity of this test is the continuity of the pdf of the underlying variable under the null hypothesis of no manipulation. Polluters and regulators have imprecise control of the waste load output because it is subject to random shocks (Brannlund and Lofgren, 1996). The pdfs of pollutant concentrations are hence expected to be continuous. The pdf of the API, on the other hand, is not continuous because it is a piece-wise linear transformation of the underlying pollutant concentrations. Figure 3 illustrates how the relationship between pollutant concentration and API has kinks at some boundaries. This is why we apply this test to the pollutant concentrations. Thus, detection of manipulation boils down to a test of whether the pdf of a pollutant concentration exhibits a discontinuity around the cut-off for blue-sky days.

In the rest of the section, we show how the piece-wise linear transformation may lead to discontinuities in the pdf of API. According to equation (1), the relationship between pollution index I and pollutant concentration C can be simplified as the following linear function:

$$I = k_1 C + k_2, \tag{3}$$

where $k_1 = (I_u - I_l)/(C_u - C_l)$ and $k_2 = I_l - C_l(I_u - I_l)/(C_u - C_l)$. Note that the values of the constant k_1 and k_2 depend on the index class of C in Table 2. In this form, pollutant concentration C is a continuous random variable and index I is piece-wise linear in C.

The cumulative distribution function (cdf) for a pollutant concentration is denoted by $F_C()$.

The cdf for the corresponding pollution index $F_I()$ is then:

$$F_I(x) = Pr\{I(C) \leq x\} = Pr\{k_1 C_{it} + k_2 \leq x\} = F_C\left(\frac{x - k_2}{k_1}\right). \tag{4}$$

Because k_1 and k_2 depend on pollution index classes, the cdf of pollution index is only differentiable piece-wisely. We can derive the pdf for the pollution index around a threshold x_c and examine the difference in probability densities between:

$$f_I(x_c^-) = \frac{1}{k_1^-} f_C\left(\frac{x^- - k_2^-}{k_1^-}\right) \text{ and } f_I(x_c^+) = \frac{1}{k_1^+} f_C\left(\frac{x^+ - k_2^+}{k_1^+}\right). \tag{5}$$

Equation (5) shows that discontinuity can occur even if the pollution data are faithfully reported. The discontinuity can be caused by the piece-wise linear transformation of pollution concentration. Let's take SO_2 as an example. The probability densities of the pollution index around 100 are $f_I(100^-) = 0.002 f_C(0.15)$ and $f_I(100^+) = 0.006 f_C(0.15)$ respectively. It is apparent that the right density is higher than the left density but this is not attributed to manipulation. The same situation is also true for SO_2. However, the pollution index for PM_{10} should be continuous around the 100 cut-off (see Figure 3). Besides that the normalized pollution index is a piece-wise linear transformation of concentration, API is the maximum of three pollutant indexes according to equation (2), which also leads to kinks at the cut-off. Therefore, applying the McCrary test to API directly will yield inconsistent results. Utilizing the confidential air quality data, we apply the discontinuity test to pollutant concentrations instead of API since the former is expected to have a continuous pdf in the absence of manipulation.

4.2 Test of Discontinuity

Since PM_{10} is the primary culprit for non-blue-sky days for the majority of Chinese cities, the main component here is the application of the McCrary test to the PM_{10} concentration. Since SO_2 and particularly NO_2 seldom cause non-blue-sky days, manipulation of these two pollutants increase the risk of being caught without improving the environmental record of local officials. It is reasonable to assume that their pollution levels should be more trustworthy. Therefore, we also apply the test to the SO_2 and NO_2 concentrations as part of our robustness checks.

The test statistic proposed by McCrary (2008) is an estimator of the log difference in height between the left and right limit of the density of the variable of interest at the cut-off, c:

$$\theta \quad = \quad \ln \lim_{r \downarrow c} f(r) - \ln \lim_{r \uparrow c} f(r) = \ln f^{+} - \ln f^{-}, \tag{6}$$

where f denotes the density of variable r, and f^{+} and f^{-} denote the right and left limit, respectively. In our case, r is the pollutant concentration and c is the cutoff for API = 100. The estimator $\hat{\theta}$ is given in the appendix. It is asymptotically normal:

$$\begin{aligned} \sqrt{nh}(\hat{\theta} - \theta) \quad \xrightarrow{d} \quad & N\left(B, \frac{24}{5}\left(\frac{1}{f^{+}} + \frac{1}{f^{-}}\right)\right) \\ & B = \frac{H}{20}\left(\frac{-f^{+"}}{f^{+}} - \frac{-f^{-"}}{f^{-}}\right), \end{aligned}$$

where h is the bandwidth and $H = \lim_{n\to\infty, h\to 0} h^2\sqrt{nh}$. We perform the one-sided lower-tailed version of the test, since we are interested in the shifting of probability mass from above the cut-off to below it, which will yield the left limit, f^{-} to be higher than the right limit, f^{+}.

The test statistic is calculated in two steps. First, using the estimated variance of the data, a bin size, b, is chosen to discretize the data and plot the first-step histogram. After that, the discretized data is used to estimate the left and right limit of the pdf at the cut-off using a bandwidth h. McCrary (2008) recommends that the ratio of bandwidth and bin size $a \equiv h/b$ shall be greater than 10. We use $a = 15$ as the benchmark. See Appendix A for the estimation process.

We use the t-statistics to infer whether a city is a manipulator. Because manipulation involves under-reporting of pollution, it causes bunching below the cut-off. Therefore, a significantly negative t-statistic is an indicator of manipulation. We also use the t-statistics to compare the degree of manipulation across cities. We rank cities by the significance of the discontinuity. The t-statistic is normalized by its variance and hence is more comparable than the actual magnitude of the discontinuity that depends on the shape of the pdf.[19] It is however important to note that a larger t-statistic does not necessarily imply a higher level of manipulation in the sense of a larger discontinuity in the pdf. Rather, it signifies a higher degree of confidence in the presence of manipulation.

[19]To make this point clear, note that for a cdf we all understand what a discontinuity of 0.1 means. However, the McCrary statistic is the difference between the logs of the left and right limit and hence is not easy to compare if the pdfs are different.

4.3 Baseline Results

We apply the McCrary test to each city using 10 years of daily pollutant concentration data. PM_{10} is the dominant pollutant in all cities, which accounts for 74% of non-blue-sky days. Hence, we expect it to be manipulated by a larger number of cities. Our baseline result, the t-statistic of the McCrary test using $a = 15$, is illustrated in Figure 4. It shows different degrees of manipulation in the PM_{10} data across cities. We categorize three levels of manipulation based on t-statistics: above -1.5 as non-manipulators, between -1.5 and -2 as borderline manipulators, and smaller than -2 as manipulators. A visual observation of this graph does not reveal obvious spatial patterns of manipulation.

The city-specific McCrary test result confirms heterogeneous manipulation behavior. Table 5 exhibits the cities that report dubious PM_{10} pollution. These cities are flagged because their pdfs of PM_{10} concentrations exhibit a statistically significant discontinuity around the cut-off. More specifically, there is a bunching below the cut-off. The baseline result, the column with $a = 15$, suggests that 61 cities reported dubious PM_{10} pollution data in the last decade. Please note that our estimation results are relatively conservative. Fifty cities show no signs of manipulation in the McCrary test.

In the introduction section, we use the API histograms of four municipalities to motivate the research question (see Figure 1). However, our formal empirical test uses pollutant concentration instead of API. The McCrary test is illustrated in Figure 5. The results confirm that Beijing has the highest likelihood of manipulation, followed by Tianjin and Chongqing. However, we cannot reject the null hypothesis that Shanghai is not a manipulator.

To summarize the McCrary test results, we plot McCrary t-statistics for the three criteria pollutants against GDP per capita, population density, and value added of the secondary industry per capita in Figure 6. These plots show that for PM_{10} the relationship between the McCrary t-statistic and the three economic variables is negative. Since a more negative McCrary t-statistic implies more significant manipulation, a negative slope implies that the higher the GDP per capita, population density, or industrial value added per capita, the more significant the manipulation of PM_{10} concentration data. This correlation is intuitive, since cities that are larger demographically or economically are more likely to have pollution problems due to PM_{10} and hence are more likely to

manipulate their data. For SO_2 and NO_2, we find that there is no correlation between the relevant McCrary t-statistics and the economic and demographic variables.

5 Patterns of Manipulation

The McCrary test does not inform under what situations cities may report dubious air pollution data. To address this issue, we propose an alternative identification strategy to investigate the patterns of manipulation. The identification relies on the fact that two cities with identical air quality should have the same API. Otherwise, the discrepancy is attributed to manipulation. This motivates our panel matching approach.

5.1 Identification Strategy

In order to detect particular patterns of manipulation, we cannot simply look at one city. The ideal experiment in this setup would be to have twin cities in the sense that they have the same distribution of true air quality. In the absence of manipulation, these twin cities, city 1 and 2, have identical distributions of API. This implies that to test whether one of the two cities manipulate, one can test the following hypothesis:

$$H_0 \quad : \quad API_{t1} \stackrel{d}{=} API_{t2}. \tag{7}$$

In this form, $\stackrel{d}{=}$ denotes equality of distribution. Since in practice we do not have twin cities, we have to form city pairs that are as close as possible in terms of true air quality.

We utilize visibility as a proxy for true air quality. Visibility measures the distance at which an object can just be discerned against a light sky (Sloane and White, 1986). Visibility degradation is closely related to air pollution because sunlight is absorbed or scattered by pollution particles in the air (Guo et al., 2009). Since visibility is reported by weather stations that are not prone to political pressures, it can serve as a reliable proxy for air pollution. However, the visibility-pollution connection is also affected by natural variations such as humidity. Therefore, we control for other weather variables including wind speed, temperature, and precipitation. These weather variables are perceived as exogenous, since they are determined by "nature" outside the economy,

or for our purposes the political economy.

Our identification strategy is described as follows. Air quality is unobservable; however, it affects both API and visibility. API is determined by both true air quality and manipulation, where the latter is the latent variable to be identified. Visibility is determined by true air quality and exogenous shocks by "nature." "Nature" is partially observable such as weather. For the unobservable "nature," we spatially cluster cities since neighboring cities share some common geographical characteristics that affect the visibility-pollution connection.

Air quality can be proxied by visibility, weather conditions, and geographical location. This allows us to compare the distributions of API of two cities in a pair on days where both face the same visibility and other weather conditions. So our revised hypothesis is:

$$H_0 \quad : \quad API_{t1}|W_{t1} = w \stackrel{d}{=} API_{t2}|W_{t2} = w. \tag{8}$$

In this form, W designates the vector of visibility and other weather variables, and w designates a value that W takes. The above hypothesis suggests that two cities have identical distributions of API conditional on visibility, weather, and geographical locations.

5.2 Why Not a Linear Model?

Now one implication of our hypothesis of no manipulation in equation (8) is that the conditional mean of API is equal for the two cities on days with the same weather variables:

$$\text{Eq. (8)} \quad \Rightarrow \quad E[API_{t1}|W_{t1} = w] = E[API_{t2}|W_{t2} = w]. \tag{9}$$

Now we can impose a linear model on the relationship between API and weather variables, as follows:

$$API_{ti} = W_{ti}'\beta + \alpha_i + \epsilon_{ti}, \tag{10}$$

for city $i = 1, 2$. In this situation, the testable implication would be:

$$\text{Eq. (8) and (10)} \quad \Rightarrow \quad \alpha_1 = \alpha_2. \tag{11}$$

Hence, the intuition behind our approach here is to test that there is no heterogeneity in the relationship between weather variables and API, once we control for geographic and provincial characteristics. For the linear model, this translates to the equality of the city-specific intercepts (city fixed effects). It is important to note that the type of manipulation that a linear model can detect is very restrictive. This is our motivation for using a very general nonparametric approach that can allow for nonlinear dependence between API and weather variables.

The nonlinear dependence between API and weather variables stems from two main sources: 1) manipulation of API, and 2) the nonlinear relationship between true air quality and weather conditions. First, the type of manipulation that can be modeled linearly only changes the mean of API. It imposes that manipulation leads the mean of API to change by the same amount whatever the weather conditions are. So it rules out a situation where manipulation occurs only under certain weather conditions. The latter is an implication of linearity. More specifically, it is due to the separability of W_{ti} and α_i in (10). We suspect that manipulation may occur without changing the mean of API. Furthermore, it is more likely to occur under weather conditions that make it harder to detect manipulation. Hence, we do not expect manipulation to be orthogonal to weather conditions. As for the second issue, the relationship between true air quality and weather conditions, especially visibility, is inherently nonlinear. For instance, visibility is a censored variable since its measure is limited to 7 or 10 miles. Hence, we expect its relationship with true air quality to be nonlinear. Given the fact that API is a nonlinear transformation of measures of true air quality, this further strengthens the argument for using a general nonlinear approach to model the relationship between API and weather variables that imposes no parametric restrictions. This is one of the primary motivations for us to use a fully nonparametric approach here.

5.3 Panel Matching Approach

Following the above identification strategy, we propose a panel matching approach to study the pattern of manipulation. Now we formalize our above discussion and show the key assumptions that lead to our hypothesis. For city pair k with cities $i = 1, 2$, we have the following relationship,[20]

$$API_{tki} \quad = \quad \xi_k(W_{tki}, \mathcal{A}_k, \mathcal{U}_{tki}), \tag{12}$$

[20]Note that (12) is not a structural relationship per se.

where API_{tki} is the API score on day t of city i which belongs to city pair k, W_{tki} are weather variables including visibility, wind speed, temperature, and precipitation, $\mathcal{A}_k$ are unobservable factors that are time-invariant and are specific to a city-pair, and $\mathcal{U}_{tki}$ represents idiosyncratic shocks. Now our key identifying assumption here is that:

$$\mathcal{U}_{tki}|W_{tk1} = w, \mathcal{A}_k = a \quad \stackrel{d}{=} \quad \mathcal{U}_{\tau k2}|W_{\tau k2} = w, \mathcal{A}_k = a, \tag{13}$$

where a is a realized value for the unobservable city-pair attribute $\mathcal{A}_k$. Please note that t is not necessarily equal to τ.[21]

Equation (13) is a homogeneity assumption, such as the one made in Chernozhukov et al. (2010) and Ghanem (2012), where similar assumptions are used to identify average partial effects in nonseparable panel models. The assumption is employed here to test the existence of manipulation. More importantly, the content of (13) is that once we control for weather conditions and unobservable factors specific to the city-pair, other unobservable factors should have the same distribution across the two cities.

Now (12) and (13) imply our testable hypothesis from above:

$$H_0 \quad : \quad API_{tk1}|W_{tk1} = w \stackrel{d}{=} API_{\tau k2}|W_{\tau k2} = w. \tag{14}$$

We are essentially matching days based on weather conditions. On days where cities 1 and 2 in pair k face the same weather conditions, their API should have the same distribution.

In order to test the equality of distribution, we apply two tests, the Kolmogorov-Smirnov test (KS test) and the t-test. The two-sample Kolmogorov-Smirnov test is a natural statistic in this setup since it detects any deviation from the equality of distributions. However, it may be conservative, when we do not have two independent samples with *i.i.d.* observations. We then run the t-test for the equality of means, which is robust to deviations from the *i.i.d.* assumptions. However, it only tests one implication of the equality of distributions, which is the equality of means.

Now for every city-pair, we test the equality of the distribution of API on days with similar

[21] This is not a restriction, because the two cities do not have to face the same weather conditions on the same day. We just need to compare days with the same weather conditions, not the same days with the same weather conditions. We do this mainly because of constraints on sample size.

weather variables including precipitation, wind speed, temperature, and visibility. We discretize our weather variables and implement the KS-test on each possible combination of weather variables.[22] One can motivate the approach here as an extension of matching methods, where we compare the distribution of API on days where the two cities face similar weather conditions. Recall that in propensity score matching, one matches individuals according to the propensity score, i.e. the predicted probability of treatment. In this paper, we match days based on weather conditions. Then, we test the equality of the API distributions of the two cities for those days.

Since we apply the same tests for all different weather combinations for every city pair, we correct for multiple testing using Romano and Shaikh (2006). For details on the specific procedure that we use, please see Appendix B.

5.4 Baseline Results

First of all, we need to form city pairs before using the panel matching method. It is implemented in two steps. First, we find nearest neighbors in terms of geographical distance for each city, and define a candidate city pair if both cities are each other's nearest neighbors.[23] Then, we remove all candidate city pairs that are not in the same province to ensure that each city pair is faced with the same provincial environmental goals. Our method results in 14 city pairs. We discard one of these pairs, Kelamayi-Wulumuqi, since the geographical distance between them is quite large. Hence, we are then left with 13 city pairs.

Tables SA.1-SA.15 in the Supplementary Appendix show the results for our panel matching approach. Among the 13 city pairs that we examine, we have four city pairs only that do not show any evidence of manipulation, specifically Wuhu-Maanshan, Zhenjiang-Yangzhou, Changzhou-Wuxi and Jilin-Changchun. We find at least some rejections for most city pairs, including Kaifeng-Zhengzhou, Quanzhou-Xiamen, Hangzhou-Shaoxing, Shenyang-Fushun, Yinchuan-Shizuishan, Xian-Xianyang, and Zhuzhou-Xiangtan.

With the exception of Xian-Xianyang, the rejections seem to occur mostly for higher levels of visibility and low wind speed. This is intuitive for two reasons. First, since poor visibility is associated with high levels of pollution, it is harder to detect manipulation. This is an important

[22] For details on discretization, see Appendix C.

[23] Nearest neighbor matching does not result in unique matches, this is why we impose this condition.

concern for the local governments because the API data are published on a daily basis and the citizens can detect whether it is reasonable to think that a particular day is a blue-sky day or not. Secondly, it is more likely for manipulation to occur when it can make a difference, i.e. when it would turn a pollution day to a blue-sky day. In addition, to make it less detectable, we are more likely to see that manipulation occurs closer to the cut-off, i.e. the pollution levels should not be that severe. This again confirms our intuition that manipulation is more likely to occur with higher levels of visibility. It is also intuitive why manipulation would occur when wind speed is low. Note that if wind speed is high, the pollutants could be "gone with the wind." However, if wind speed is low, then nature is not helping reduce pollution. As a result, cities manipulate their API to meet the target.

It is important to note here that the fact that we find manipulation occurs for certain weather conditions but not others indicates that a linear specification is not appropriate. Recall from above, that a linear specification implies that manipulation is orthogonal to the weather conditions. Our results indicate that this is not the case. Furthermore, they illustrate that the measurement error resulting from manipulation is expected to violate the classical measurement-error assumptions.

To illustrate the formal results in Tables SA.1-SA.15 of the Supplementary Appendix and to gain some intuition for the approach, we also include Figures SA.1-SA.13. Figure 7 is an example of illustration for the city pair Zhejiang and Yangzhou. Each figure contains 6 plots for different weather variable combinations for each city pair. VSB denotes visibility, WSP wind speed, and TEMP denotes temperature. All figures are for precipitation equal to zero after discretization, see Appendix C. Each plot includes two empirical cumulative distribution functions (cdfs) in solid lines, one for each city in a pair. We also include point-wise 95% confidence bands for each empirical cdf using dotted lines. It is important to note that these confidence bands are only to be used as a reference point and may not be interpreted formally, since the significance level that is used for the formal test is adjusted to correct for multiple testing. Furthermore, since we are comparing two functions, we ought to use uniform confidence bands, which would be wider than the point-wise ones. Hence, caution is needed in interpreting the point-wise confidence bands.

Figures SA.1, SA.2, SA.3, and SA.13 show how the city pairs Zhenjiang-Yangzhou, Changzhou-Wuxi, Jilin-Changchun, and Wuhu-Maanshan, respectively, do not reflect manipulation under various weather conditions. This of course gives us confidence in our approach that it can detect

the absence of manipulation.

For the other city pairs, we do find manipulation, but what the figures show is that different city pairs manipulate in different ways. In most cases, the empirical cdf of API of the non-manipulating city first-order stochastically dominates that of the manipulating city. The reasoning here is that the manipulating city will tend to shift mass to the lower values of API throughout the entire distribution. For higher levels of visibility, this occurs for Kaifeng-Zhengzhou, Zhuzhou-Xiangtan, Quanzhou-Xiamen, Hangzhou-Shaoxing, Shenyang-Fushun, Jinan-Taian, and Huhehaote-Baotou. It is also important to note that the degree of manipulation, which is represented graphically by the vertical distance between the two cdfs, may be very different according to the weather conditions. For instance, for Huhehaote-Baotou, when wind speed and temperature is low, manipulation is much more severe than with higher temperature and wind speed.

The manipulation that occurs for city pair Xian-Xianyang, however, does not lead to first-order stochastic dominance. The upper-middle plot in Figure SA.10 shows that the empirical cdf of Xian is flat between 100 and 125, which is clear evidence of manipulation. For this case, it seems that manipulation occurs mostly right around the cut-off.

6 Robustness Checks and Discussion

Since we are dealing with observational data, we have to check the robustness of our approach and also discuss potential caveats that may affect how we interpret our results. This section is devoted to this end.

6.1 Robustness Checks

First, we investigate the impact of a, the ratio of bandwidth to bin size, on the McCrary test results. According to McCrary (2008), the choice of bin size has no consequences on the test statistic if $a > 10$. Now in order to ensure that the asymptotic approximation delivers correct inference in finite samples, we need h to be fairly small, hence we choose to rank the cities' statistics using $a = 15$. In the robustness checks, we allow different values for a. The McCrary test results with $a = 10$ or 20 are reported alongside with the baseline results in Tables 5 - 7. The choice of bandwidth and bin size matters in the test results. It affects not only the significance but also the ranking of the cities

in terms of the likelihood of manipulation. The results show that the larger the bandwidth or the smaller the bin size the more likely it is a city to be tagged as a manipulator. This is a standard issue in nonparametric methods.

The second robustness check is concerned with the manipulation of different pollutants. Since PM_{10} caused 73.7% of the total pollution days, it is the most vulnerable target. We find that 55% of cities reported dubious PM_{10} data. SO_2 and NO_2 account for 9.2% and 0.2% of primary pollutants respectively. There should be less cities manipulating SO_2 and NO_2 concentrations. We report the results for the SO_2 concentrations in Table 6. We find that , 26 cities, or 23%, are flagged because of the bunching below the cut-off. As for NO_2, we include its results in the Supplementary Appendix. NO_2 seldom leads to API greater than the cut-off for blue-sky days. Hence, there is very little mass above that cut-off to be able to estimate the right limit of the pdf, which makes the results of the McCrary test unreliable.

The third robustness check involves implementing the McCrary test for "artificial" cut-offs. We change the cut-off to $c = 0.1$ and $c = 0.2$ in lieu of the cut-off for blue-sky day. If our hypothesis is true that manipulation leads to a discontinuity, then we do not expect to find any rejection for the artificial cut-offs. We implement the test for both PM_{10} and SO_2 and report the results in the Supplementary Appendix. We do not find evidence of a discontinuity for either of the pollutants.[24]

The fourth robustness check compares the application of the McCrary test to pollutant concentrations and API. We have demonstrated in the identification section that applying the test to API directly might lead to inconsistent estimates because some discontinuities in the API distribution are inevitable because it is a piece-wise linear transformation of pollutant concentration. In order to empirically illustrate this argument, we also apply our test to API and report the implications for our key results in Table 8. Because there is no kink in the transformation from PM_{10} to API at API = 100, see Figure 3, the McCrary test shall yield similar results for the concentration and API. The major difference is in SO_2. The concentration model suggests 19 manipulators while the API model only identifies 4 manipulators. The results on NO_2 are close and we attribute it to the fact that the number of observations is small because NO_2 seldom showed up as a primary pollutant.

Fifth, as a robustness check for the panel matching approach, we compare its result with the

[24]There are very few significant results that are driven by little mass above the cut-off which leads to unreliable estimates of the right limit of the pdf.

McCrary result in Table 9. "YES" implies that the relevant test finds evidence for manipulation, "NO" implies the contrary, and "Borderline" implies that it depends on the bandwidth. If the panel matching approach finds rejections, then this may indicate that either one city in the pair is manipulating or both cities manipulate in different ways. If we do not find rejections, then most likely both cities should not be manipulating. It is however possible, though unlikely, that those cities manipulate in exactly the same way. Finally, we may find manipulation in the panel matching approach but not in the McCrary test. This would be the case, if manipulation behavior does not lead to a discontinuity in the pdf of the pollutant concentrations at the cut-off.

As we expect, we do not find a city pair, where both cities manipulate according to the McCrary test but we find no rejections in the panel matching approach. However, we find three pairs, where only one city in the pair manipulates according to the McCrary test, but our panel matching approach finds no rejections. This is the case for Zhenjiang-Yangzhou, Changzhou-Wuxi and Jilin-Changchun. This may be due to the relatively small subsample size for these three city-pairs. It also may be because the panel matching approach does not test for discontinuities so by construction it is less powerful at detecting this form of manipulation. This phenomenon may be explained from a political-economy perspective. For instance, it is possible that a city manipulates in a way to keep up with neighboring cities, if local government officials in these cities compete over promotions. It is also important to point out that the KS statistic may be conservative in our setting, since it may violate the classical assumptions under which the KS asymptotic distribution is derived. Since we have a time-series cross-section, there may be time-series dependence that would render the KS statistic conservative.

For the rest of the city-pairs, we find that our panel matching approach and the McCrary test yield consistent results. This is in line with our expectations that both approaches should confirm each other once we exclude some unlikely cases. As noted above, if there is manipulation that does not lead to a discontinuity at the cut-off, the McCrary test cannot detect it. However, the panel matching approach can detect manipulation regardless of the presence of a discontinuity. This is exactly the case for the Hangzhou-Shaoxing city pair. The McCrary test does not indicate manipulation for either city, yet we find evidence for manipulation in the panel matching approach. It is important to note here that for manipulation to lead to a discontinuity in the pdf of the pollutant concentrations there needs to be very strong manipulation around the cut-off. Since

both of these cities are coastal, it is conceivable that they only manipulate their data slightly such that one cannot detect a discontinuity in the pdf of their pollutant concentrations.

6.2 Further Discussion

It is important to stress that our empirical results are suggestive. What we refer to as manipulation may not be actual manipulation if our assumptions do not hold. For the McCrary test, the main caveat is that one can only detect the types of manipulation that lead to a discontinuity. For instance, if manipulation leads to a mean shift in the distribution of the concentration, then the test has no power against this alternative by construction. There are other ways of manipulation that may not result in a discontinuity, but we do not find these alternatives very likely in practice. For cities to manipulate without leading to a discontinuity at the cut-off for blue-sky days, they must have knowledge of the distribution of the concentration for the entire period that we are studying. However, cities have to report their data on a daily basis and hence it is rather unlikely that they can manipulate without leading to a discontinuity at the cut-off.

Discontinuity is an alarming signal but clustering of pollutant distribution around the cutoff is not necessarily due to manipulation. Risk-averse polluters might over-comply with the standards and cause bunching of pollution data (Bandyopadhyay and Horowitz, 2006; Earnhart, 2007; Shimshack and Ward, 2008). In our case, in order to achieve a certain number of blue-sky days, some local governments temporarily shut down factories, reduce energy supply, or require firms to use high-quality fuels.[25] These activities will also cause clustered pollutant distributions. Although these short-run policies are not efficient, we cannot label these local officials as manipulators. However, as long as air quality cannot be precisely controlled, clustering should not lead to a discontinuity. Therefore, our discontinuity test is still a valid means of detecting manipulation. In addition, clustering does not affect the relationship between API, visibility and other weather variables. In this case, our panel matching approach is also robust.

The panel matching approach uses the KS test, which does not directly identify which city in a pair manipulates. We do not think this is a downside to this approach, since our objective here is to learn about patterns of manipulation rather than detecting which city is a manipulator. The McCrary test is more geared toward pointing out manipulators. Hence, we see the two approaches

[25] Personal communication with the officials from the Ministry of Environmental Protection of China.

as complements rather than substitutes.

The main caveat of the panel matching approach is that it relies heavily on the assumption that once we condition on weather variables, the *entire* distribution of API should be the same for both cities. This is of course a strong assumption. If we thought that there is still a geographical difference between the two cities that may lead their distribution of API to be different under certain weather conditions, such as high visibility, but not otherwise, then this would confound the results of our panel matching approach. However, the manipulation that we find using the panel matching approach is consistent with the results from the McCrary test. Even though the formal tests of the panel matching approach do not indicate which city manipulates, examining the figures of the empirical cdfs can suggest which city is the manipulator as we discussed above. For instance, Figure SA.4 shows that Kaifeng is the one that moves mass toward the lower values, hence suggesting that it is the manipulator. This is consistent with the McCrary test results. Similar examination of Figures SA.5, SA.6, SA.8 and SA.10 show that the panel matching approach is detecting manipulation.

For the pair of Hangzhou-Shaoxing, we do find very few instances of manipulation using the panel matching approach, even though the McCrary test does not detect manipulation for either city. This is not surprising, because if manipulation is mild, then it may not lead to a discontinuity in the unconditional distribution of the pollutant concentration. Hence, it cannot be detected by the McCrary test, but may be detected by the panel matching approach.

7 Conclusion

We find that the daily air pollution data in China are not well-behaved. The assertion is based on the analysis that applies a discontinuity test and a panel matching approach to a unique data set that covers all major cities over a decade. These results are relevant for empirical researchers and policy makers who may use such data to learn about the effects of pollution on various types of outcomes. We suggest that thorough robustness checks need to be done to examine the impact of the cut-offs. Particular attention should be paid to the cities that are on our suspect list. In addition, we find a fair amount of heterogeneity and non-linearity in the data reporting behavior. Thus, the resulting data are unlikely to reflect the classical measurement error assumptions. Our

results indicate that the use of standard methods that rely on the classical measurement-error assumptions would not be appropriate. Hence, the use of such data requires caution and care in the choice of estimation strategy and assumptions on the measurement error.

Our methodology can help the monitors ferret out the cities that report dubious data. In particular, we have discovered the meteorological conditions under which local officials are more likely to manipulate. However, the conviction of a manipulator requires an independent direct measure of pollution because ours only provides suggestive evidence. Our results suggest that situations where government officials report data that are used in their own performance evaluation lead to strong incentives for manipulation, as we would expect. Therefore, our models are applicable not only to daily air pollution reports but also to other self-reported data.

This paper has focused on the identification of potential manipulation behavior. Although we have done some preliminary analyses on the patterns of manipulation, we did not provide a political economy interpretation why manipulation is more likely to occur in some cities but not others. This question will be left for future studies.

A McCrary Test Statistic

A.1 Implementation given b and h

There are two steps to implementing the McCrary (2008) test statistic on a variable Z_i:

1. First-step Histogram of the Discretized X_i

   Using a binsize b,

$$g(Z_i) = \lfloor \frac{Z_i - c}{b} \rfloor b + \frac{b}{2} + c \in \left\{ ..., c - 5\frac{b}{2}, c - 3\frac{b}{2}, c - \frac{b}{2}, c + \frac{b}{2}, c + 5\frac{b}{2}, ... \right\} \tag{15}$$

   where $\lfloor a \rfloor$ is the greatest integer function.

   Now $\{X_j\}_{j=1}^J$ is the equi-spaced grid of width b covering the support of $g(Z_i)$ and

$$Y_j = \frac{1}{nb} \sum_{i=1}^{n} 1\{g(R_i) = X_j\} \tag{16}$$

   A scatter plot of X_j and Y_j gives the first-step histogram.

2. Calculation of $\hat{\theta}$

$$\begin{aligned} \hat{\theta} &= \ln \hat{f}^+ - \ln \hat{f}^- \\ &= \ln \left\{ \sum_{X_j > c} K\left(\frac{X_j - c}{h}\right) \frac{S^+_{n,2} - S^+_{n,1}(X_j - c)}{S^+_{n,2} S^+_{n,0} - (S^+_{n,1})^2} Y_j \right\} \\ &\quad - \ln \left\{ \sum_{X_j < c} K\left(\frac{X_j - c}{h}\right) \frac{S^-_{n,2} - S^-_{n,1}(X_j - c)}{S^-_{n,2} S^-_{n,0} - (S^-_{n,1})^2} Y_j \right\}, \end{aligned}$$

where $S^+_{n,k} = \sum_{X_j > c} K((X_j - c)/h)(X_j - c)^k$, $S^-_{n,k} = \sum_{X_j < c} K((X_j - c)/h)(X_j - c)^k$, and $K(t) = max\{0, 1 - |t|\}$.

A.2 Selection of b and h

McCrary (2008) points out that the binsize does not matter provided that $h/b > 10$. We use $\hat{b} = 2\hat{\sigma} n^{-1/2}$ proposed in the bandwidth selection guide, where $\hat{\sigma}$ is the standard deviation of Z_i. In terms of bandwidth selection, McCrary (2008) proposes a method of bandwidth selection

based on the rule of the thumb of Fan and Gijbels (1996). The method is based on a global 4th order polynomial approximation on either side of the cut-off. Unfortunately, the regression is sparse for our case. Hence we choose $h = a * b$ where $a \in \{10, 15, 20\}$. We use relatively small bandwidths, since the normal approximation behaves poorly when the bandwidth is large due to the bias term.[26]

A.3 Local Linear Estimator for Density Plots

The density plots are local linear estimators of the density on the right and left of the cut-off, as follows

$$\begin{aligned}(\hat{\phi}_1, \hat{\phi}_2) &\equiv \arg\min L(\phi_1, \phi_2, r) \\ &= \sum_{j=1}^{J} \{Y_j - \phi_1 - \phi_2(X_j - r)\}^2 K\left(\frac{X_j - r}{h}\right) \\ &\quad (1\{X_j > c\}1\{r \geq c\} + 1\{X_j < c\}1\{r < c\})\end{aligned}$$

B Correction for Multiple Testing

To correct for multiple testing, for each city pair we use a nominal value of $\alpha = 0.05$ and then apply Romano and Shaikh (2006) step-up procedure to control the k-FWER with $k = 1$ and the initial α_i as defined in (13) in Romano and Shaikh (2006).

$$\alpha_i = \begin{cases} \frac{k}{s}, & \text{if} i \leq k, \\ \frac{k}{s+k-i}, & \text{if} i > k, \end{cases}, \tag{17}$$

where s denotes the total number of hypotheses, $H_1, H_2, ..., H_s$, and for our case this is equivalent to the total number of weather variable combinations. Together with $D_1(k, s)$ defined as

$$\begin{aligned}D_1(k, s) &= \max_{k \leq |I| \leq s} S_1(k, s, |I|), \\ S_1(k, s, |I|) &= |I| \frac{\alpha_{s-|I|+k}}{k} + |I| \sum_{k<j<|I|} \frac{\alpha_{s-|I|+j} - \alpha_{s-|I|+j-1}}{j},\end{aligned} \tag{18}$$

[26] We may want to estimate bias to double-check that our results are not affected by this issue.

we can construct the critical values for our p-values, which we denote by p_i^c,[27]

$$p_i^c = \frac{\alpha\alpha_i'}{D_1(k,s)} \tag{19}$$

To apply the step-up procedure, we order the observed p-values $\hat{p}_i$ in an ascending order, and let $\hat{p}_{(i)}$ denote the i^{th} smallest p-value. Now the step-up procedure start with the largest p-value, $\hat{p}_{(s)}$. If $\hat{p}_{(s)} \leq p_s^c$, then reject all hypotheses $H_1, ..., H_s$. Otherwise, reject $H_1, ..., H_j$, such that j is the smallest integer for which,

$$\hat{p}_{(s)} > p_s^c, ..., \hat{p}_{(j+1)}^c > p_{j+1}^c. \tag{20}$$

C Discretization of Weather Variables

Visibility is rounded to the nearest integer. Precipitation is divided into three equi-space bins $[0, 0.33]$, $(0.33, 0.67]$ and $(0.67, 1]$, which we refer to in the table by 0, 0.5 and 1, respectively. Wind speed is divided into 5 bins, $[0, 5]$, $(5, 10]$, $(10, 15]$, $(15, 20]$, $(20, 25]$. Finally, temperature is also divided into 5 bins ≤ 20, $(20, 40]$, $(40, 60]$, $(60, 80]$, and $(80, 100]$.

[27] Romano and Shaikh (2006) denote it by $\tilde{\alpha}_i$.

References

Andrews, Steven Q. 2008a. "Beijing Plays Air Quality Games." *Far Eastern Economic Review* .

Andrews, Steven Q. 2008b. "Inconsistencies in air quality metrics: 'Blue Sky' days and PM 10 concentrations in Beijing." *Environmental Research Letters* 3 (3):034009.

Bandyopadhyay, Sushenjit and John Horowitz. 2006. "Do Plants Overcomply with Water Pollution Regulations? The Role of Discharge Variability." *The B.E. Journal of Economic Analysis & Policy (Topics)* 6 (1):4.

Brannlund, Runar and Karl-Gustaf Lofgren. 1996. "Emission Standards and Stochastic Waste Load." *Land Economics* 72 (2):218–230.

Chay, Kenneth Y. and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy* 113 (2):pp. 376–424.

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 – 425.

Chen, Yuyu, Ginger Zhe Jin, Naresh Kumar, and Guang Shi. 2012. "Gaming in Air Pollution Data? Lessons from China." *The B.E. Journal of Economic Analysis & Policy (Advances)* 13 (3):Article 2.

Chernozhukov, Victor, Ivan Fernandez-Val, Jinyong Hahn, and Whitney Newey. 2010. "Average and Quantile Effects in Nonseparable Panel Data Models." *Unpublished Manuscript* .

Congleton, Roger D. 1992. "Political Institutions and Pollution Control." *The Review of Economics and Statistics* 74 (3):412–21.

Earnhart, Dietrich. 2007. "Effects of permitted effluent limits on environmental compliance levels." *Ecological Economics* 61 (1):178 – 193.

Fan, J. and I. Gijbels. 1996. "Local Polynomial Modelling and its Applications." *Unpublished Manuscript* .

Ghanem, Dalia. 2012. "Average Partial Effects in Nonseparable Panel Data Models: Identification and Testing." *Unpublished Manuscript* .

Gibbons, Charles, Juan Carlos Serrato, and Michael Urbancic. 2011. "Broken or Fixed Effects?" *Unpublished Manuscript* .

Graff Zivin, Joshua S. and Matthew J. Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review* Forthcoming (17004).

Guo, Jian-Ping, Xiao-Ye Zhang, Hui-Zheng Che, Sun-Ling Gong, Xingqin An, Chun-Xiang Cao, Jie Guang, Hao Zhang, Ya-Qiang Wang, Xiao-Chun Zhang, Min Xue, and Xiao-Wen Li. 2009. "Correlation between PM concentrations and aerosol optical depth in eastern China." *Atmospheric Environment* 43 (37):5876 – 5886.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *The Quarterly Journal of Economics* 125 (1):307–362.

Koech, Janet and Jian Wang. 2012. "Chinas Slowdown May Be Worse Than Official Data Suggest." *Federal Reserve Bank of Dallas Economic Letter* 7 (8):1–4.

Koop, Gary and Lise Tole. 2004. "Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air?" *Journal of Environmental Economics and Management* 47 (1):30 – 54.

Li, Hongbin and Li-An Zhou. 2005. "Political turnover and economic performance: the incentive role of personnel control in China." *Journal of Public Economics* 89 (910):1743 – 1762.

Malm, William C. 1999. *Introduction to Visibility*. Fort Collins, CO: Cooperative Institute for Research in the Atmosphere (CIRA), NPS Visibility Program, Colorado State University.

Matus, Kira, Kyung-Min Nam, Noelle E. Selin, Lok N. Lamsal, John M. Reilly, and Sergey Paltsev. 2012. "Health damages from air pollution in China." *Global Environmental Change* 22 (1):55 – 66.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142 (2):698–714.

Natural Resources Defense Council. 2009. "Amending China's Air Pollution Prevention And Control Law." Available at http://www.nrdc.cn/.

Rawski, Thomas G. 2001. "What is happening to China's GDP statistics?" *China Economic Review* 12 (4):347 – 354.

Romano, Joseph P. and Azeem M. Shaikh. 2006. "Stepup procedures for control of generalizations of the familywise error rate." *Annals of Statistics* 34 (4):1850–1873.

Shih, Victor, Christopher Adolph, and Mingxing Liu. 2012. "Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China." *American Political Science Review* 106:166–187.

Shimshack, Jay P. and Michael B. Ward. 2008. "Enforcement and over-compliance." *Journal of Environmental Economics and Management* 55 (1):90 – 105.

Sloane, Christine S. and Warren H. White. 1986. "Visibility: An evolving issue." *Environmental Science & Technology* 20 (8):760–766.

World Bank. 2007. "Cost of Pollution in China." World Bank, East Asia and Pacific Region, Available at http://www.worldbank.org/eapenvironment.

Zhang, Junjie. 2012. "Delivering Environmentally Sustainable Economic Growth: The Case of China." Asia Society Policy Report, Available at http://asiasociety.org/policy/.

Zhang, Junjie and Can Wang. 2011. "Co-benefits and additionality of the clean development mechanism: An empirical analysis." *Journal of Environmental Economics and Management* 62 (2):140 – 154.

Zhang, Qingfeng and Robert Crooks. 2012. *Toward an Environmentally Sustainable Future: Country Environmental Analysis of the Peoples Republic of China*. Mandaluyong City, Philippines: Asian Development Bank.

Zheng, Siqi, Matthew E. Kahn, Weizeng Sun, and Danglun Luo. 2013. "Incentivizing Chinas Urban Mayors to Mitigate Pollution Externalities: The Role of the Central Government and Public Environmentalism." NBER Working Papers 18872, National Bureau of Economic Research, Inc.

Table 1: Ambient air quality standards

Year	Standard	Criteria Air Pollutants
1982	GB3905-1982	TSP, Airborne dust, SO_2, NO_x, O_3, CO
1996	GB3905-1996	SO_2, TSP, PM_{10}, NO_x, NO_2, CO, O_3, Pb, BaP, F
2000	GB3905-1996 (Revision)	SO_2, TSP, PM_{10}, NO_2, CO, O_3, Pb, BaP, F
2012	GB3905-2012	SO_2, NO_2, CO, O_3, PM_{10}, $PM_{2.5}$, TSP, NO_x, Pb, BaP

Table 2: Air pollution index and corresponding pollution levels

	Daily Average Concentrations (mg/m^3)		
API	Sulfur Dioxide (SO_2)	Nitrogen Dioxide (NO_2)	Particulate Matter (PM_{10})
50	0.050	0.080	0.050
100	0.150	0.120	0.150
200	0.800	0.280	0.350
300	1.600	0.565	0.420
400	2.100	0.750	0.500
500	2.620	0.940	0.600

Table 3: Categories of API and health concerns

API	Air Quality Level	Air Quality Conditions	Health Concerns
0-50	I	Excellent	Good
51-100	II	Good	Moderate
101-200	III	Lightly polluted	Unhealthy for sensitive groups
201-300	IV	Moderately polluted	Unhealthy
>300	V	Heavily polluted	Hazardous

Table 4: Summary statistics of the variables used in the analysis

	Mean	Std	Min	Max
API (max)	76.315	37.680	0	625
Primary Pollutant				
SO_2	0.092			
NO_2	0.002			
PM_{10}	0.737			
Grade				
I	0.169			
II	0.677			
III1	0.124			
III2	0.021			
IV1	0.003			
IV2	0.002			
V	0.004			
Pollution Concentration (mg/m^3)				
SO_2	0.054	0.055	0.001	2.147
NO_2	0.036	0.020	0.001	0.353
PM_{10}	0.100	0.066	0.001	2.721
API (pollutant level)				
SO_2	43.262	27.544	0	409
NO_2	22.850	14.116	0	226
PM_{10}	74.731	38.277	0	501
Weather				
Visibility (VSB)	6.497	2.541	0	10
Wind speed (SPD)	5.289	2.886	0	42.75
Temperature (TEMP)	59.081	19.604	-23.50	98.04
Pressure (STP)	965.954	76.455	637.78	1045.48
Precipitation (PCP)	0.069	0.263	0	11.30

Table 5: t-Statistics for McCrary Test: Cities Exhibiting Manipulation (PM_{10})

	a				a		
	10	15	20		10	15	20
Shenyang	-10.318	-12.514	-14.091	Quanzhou	-3.012	-3.414	-4.043
Shijiazhuang	-8.620	-10.609	-11.865	Zaozhuang	-1.700	-3.330	-3.975
Anshan	-7.612	-8.748	-9.528	Nantong	-2.285	-3.160	-3.145
Beijing	-7.800	-8.673	-9.064	Dalian	-2.372	-2.998	-3.742
Chengdu	-6.407	-8.588	-10.079	Changzhi	-2.548	-2.992	-3.156
Nanjing	-7.291	-8.277	-8.862	Anyang	-3.075	-2.992	-3.161
Hefei	-6.648	-8.275	-9.239	Lanzhou	-2.069	-2.931	-4.144
Hangzhou	-6.235	-7.735	-8.441	Weifang	-0.821	-2.833	-4.159
Xining	-5.286	-6.294	-6.984	Datong	-2.935	-2.803	-2.693
Weinan	-4.828	-6.144	-6.899	Jinan	-1.356	-2.757	-3.841
Changchun	-3.602	-5.541	-6.750	Zhenjiang	-1.512	-2.697	-3.553
Luoyang	-5.198	-5.409	-5.307	Guiyang	-2.226	-2.413	-2.327
Chongqing	-4.167	-5.307	-5.761	Yinchuan	-1.806	-2.045	-2.541
Jinzhou	-4.054	-5.275	-5.728	Baoji	-1.367	-1.979	-2.590
Tianjin	-3.759	-4.725	-5.510	Deyang	-1.063	-1.975	-2.406
Kaifeng	-3.604	-4.663	-5.622	Xianyang	-0.467	-1.930	-2.966
Xiangtan	-3.514	-4.649	-5.437	Yanan	-1.282	-1.882	-2.202
Qingdao	-3.300	-4.625	-4.985	Rizhao	-1.545	-1.841	-2.214
Suzhou	-3.066	-4.609	-5.728	Mianyang	-1.671	-1.803	-1.935
Xuzhou	-3.301	-4.133	-4.539	Lianyungang	-0.757	-1.738	-2.194
Guiyang	-3.072	-3.982	-4.812	Yangquan	-1.731	-1.736	-1.524
Changzhou	-3.574	-3.905	-4.079	Shenzhen	-0.572	-1.656	-2.540
Baotou	-2.856	-3.829	-4.532	Nanchang	-0.858	-1.633	-2.358
Taiyuan	-2.833	-3.820	-5.254	Tongchuan	-0.581	-1.576	-1.859
Linfen	-3.466	-3.803	-4.056	Luzhou	-1.671	-1.620	-1.278
Xian	-1.848	-3.672	-4.914	Taian	-0.681	-1.572	-2.472
Wulumuqi	-3.582	-3.672	-4.269	Yantai	-0.873	-1.288	-1.879
Wuhan	-3.242	-3.534	-3.355	Yangzhou	-0.673	-1.237	-1.663
Guangzhou	-3.113	-3.497	-3.324	Lasa	-0.711	-1.099	-1.678
Jining	-2.109	-3.475	-3.973	Shaoxing	-1.030	-1.544	-1.463
				Huhehaote	-1.066	-1.325	-1.614

This table reports the McCrary t-statistic for each city using the PM_{10} concentration data.
Notes: $a = h/\hat{b}$, h is the bandwidth and $\hat{b} = 2\hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is the standard deviation of the pollutant concentration. For more details, please see Appendix A.

Table 6: t-Statistics for McCrary Test: Cities Exhibiting Manipulation of SO_2

	a		
	10	15	20
Shijiazhuang	-4.9154	-5.4971	-5.7896
Anshan	-4.3499	-5.1613	-5.5384
Yangquan	-3.7741	-4.7820	-5.4176
Zunyi	-3.1786	-4.6268	-5.1011
Tangshan	-2.8904	-3.8594	-4.2408
Linfen	-2.7670	-3.5816	-3.1446
Weifang	-2.1114	-3.1647	-4.0624
Handan	-3.4122	-3.0500	-2.6932
Shizuishan	-2.2226	-2.9100	-3.5842
Taiyuan	-1.2444	-2.7966	-3.9241
Dalian	-2.7057	-2.6835	-2.6973
Tianjin	-1.6428	-2.4047	-2.3984
Zhengzhou	-1.4686	-2.1941	-2.6340
Kunming	-1.9936	-2.1755	-2.0017
Guiyang	-1.5683	-2.0856	-2.5199
Baoding	-1.3858	-1.8321	-2.2295
Jinzhou	-1.3630	-1.7740	-1.7233
Kaifeng	-0.5515	-1.7361	-1.6730
Baotou	-1.4601	-1.6942	-1.7832
Wulumuqi	-1.3142	-1.5663	-1.9597
Changzhi	-1.0818	-1.4499	-1.6063
Shenyang	-0.3135	-1.3845	-2.1399
Jinan	-0.9440	-1.3724	-1.2364
Zaozhuang	-1.3397	-1.3693	-1.3118
Datong	-1.3776	-1.3352	-0.4627
Weinan	-0.4897	-1.3234	-2.0854

This table reports the McCrary t-statistic for each city using the SO_2 concentration data.

Notes: $a = h/\hat{b}$, h is the bandwidth and $\hat{b} = 2\hat{\sigma}/\sqrt{n}$ where $\hat{\sigma}$ of the pollutant concentration. For more details, please see Appendix A.

Table 7: t-Statistics for McCrary Test: NO_2

	a				a		
	10	15	20		10	15	2
Guangzhou	-3.516	-3.409	-3.824	Hangzhou	1.034	0.115	-0.664
Mudanjiang	-3.237	-3.222	-3.152	Xianyang	NaN	0.150	0.379
Shenzhen	-2.502	-3.145	-3.678	Weinan	0.996	0.174	-0.283
Luoyang	-1.666	-2.247	-2.237	Changzhi	0.034	0.360	0.388
Yinchuan	NaN	-1.741	-6.336	Lanzhou	0.795	0.409	0.046
Anyang	-1.850	-1.562	-1.451	Shijiazhuang	1.074	0.418	0.657
Kaifeng	-1.392	-1.402	-1.276	Xiamen	-2.523	0.475	0.259
Nanning	NaN	-1.318	-1.252	Nanjing	0.987	0.547	0.263
Beijing	-1.425	-1.314	-1.378	Wuhan	0.848	0.558	0.309
Fushun	NaN	-1.207	-1.300	Linfen	0.311	0.741	0.474
Ningbo	-0.929	-1.140	-1.275	Tianjin	0.524	0.871	1.029
Haerbin	-1.775	-1.104	-0.692	Chongqing	1.067	0.921	0.643
Wuxi	NaN	-1.000	-2.413	Huhehaote	0.890	1.121	1.440
Huzhou	-0.518	-0.732	-0.741	Yangquan	NaN	1.135	0.789
Xuzhou	NaN	-0.357	Inf	Datong	0.366	1.200	1.326
Wenzhou	0.217	-0.351	-0.248	Suzhou	0.717	1.213	1.659
Zhenjiang	NaN	-0.042	NaN	Baoding	1.300	1.255	1.381
Wuhu	NaN	-0.042	0.000	Zigong	1.401	1.324	1.375
Zunyi	NaN	0.000	0.000	Jiaozuo	1.071	1.502	1.398
Pingdingshan	0.000	0.000	0.281	Yanan	0.371	1.730	2.198
Yueyang	0.000	0.000	1.158	Shanghai	2.068	1.790	1.589
Benxi	0.000	0.000	-0.595	Wulumuqi	1.278	1.799	1.843
Shenyang	0.000	0.000	0.000	Handan	1.496	1.883	2.164
Changde	NaN	0.000	0.000	Nantong	-4.747	1.939	2.249
Panzhihua	NaN	0.005	0.846				

This table reports the McCrary t-statistic for each city using the NO_2 concentration data.
Notes: $a = h/\hat{b}$, h is the bandwidth and $\hat{b} = 2\hat{\sigma}/\sqrt{n}$ where $\hat{\sigma}$ of the pollutant concentration. For more details, please see Appendix A.

Table 8: McCrary test: Comparison between Concentration and API Results

	Concentration	API
No. of Manipulators ($\alpha = 0.05$)		
PM_{10}	52	50
SO_2	19	4
NO_2	5	7
Ranking (PM_{10})		
Beijing	4	7
Chongqing	13	15
Shanghai	91	95
Tianjin	15	26

This table reports the difference in the results between using pollutant concentration versus API in the McCrary test.

Table 9: PMA-McCrary Comparison

City-pair	PMA	McCrary (2008)	
	Rejections	City 1	City 2
Zhenjiang-Yangzhou	NO	Borderline	NO
Changzhou-Wuxi	NO	YES	NO
Jilin-Changchun	NO	NO	YES
Kaifeng-Zhengzhou	YES	YES	NO
Zhuzhou-Xiangtan	YES	NO	YES
Quanzhou-Xiamen	YES	YES	NO
Hangzhou-Shaoxing	YES	NO	NO
Shenyang-Fushun	YES	YES	NO
Yinchuan-Shizuishan	YES	Borderline	NO
Xian-Xianyang	YES	YES	Borderline
Jinan-Taian	YES	YES	NO
Huhehaote-Baotou	YES	NO	YES
Wuhu-Maanshan	NO	NO	NO

The table compares the results from the panel matching approach (PMA) with the McCrary (2008) test. *YES* indicates that there was evidence of manipulation, *NO* indicates that there was no evidence of manipulation, and *Borderline* indicates that there is borderline evidence of manipulation. For instance, it was a borderline p-value at the 5% level.

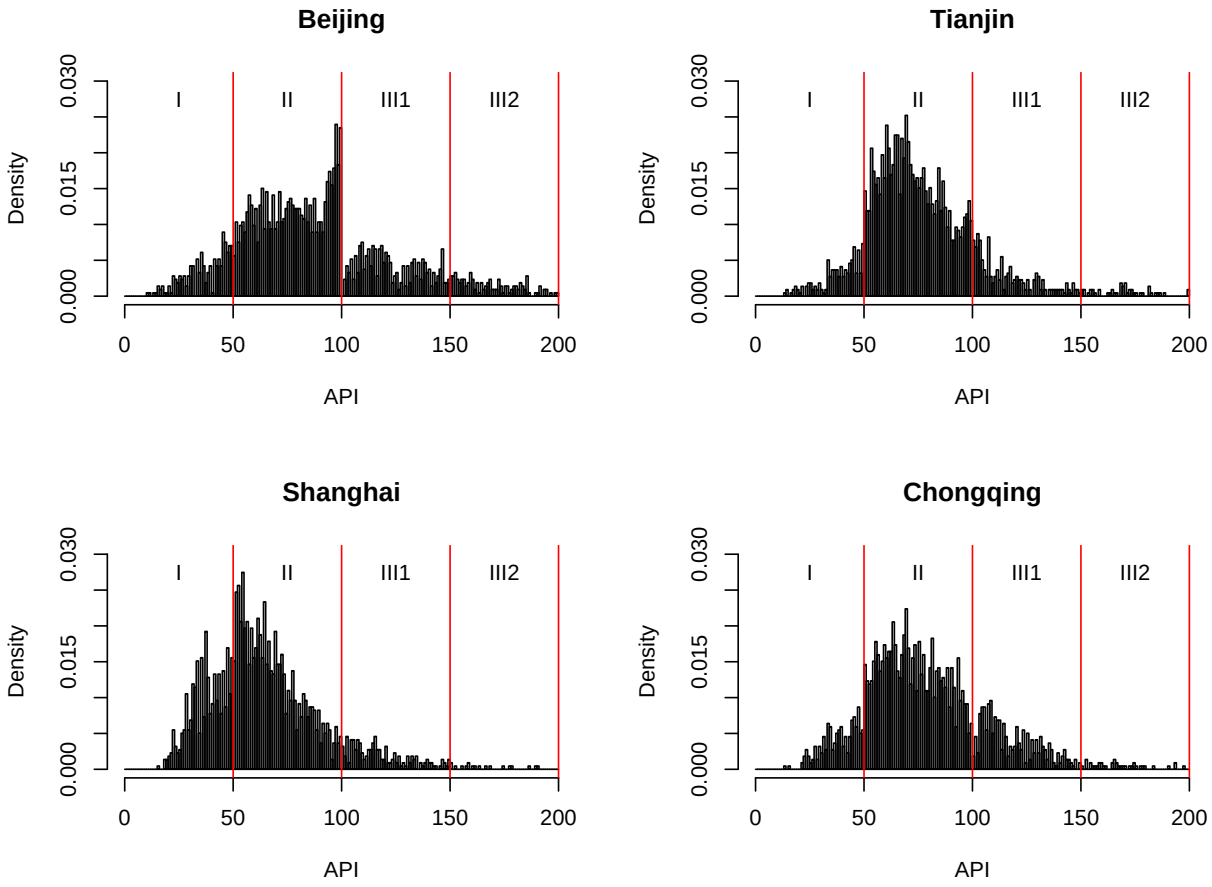


Figure 1: Frequency of API scores in four capital cities: Beijing (upper left), Tianjin (upper right), Shanghai (lower left), and Chongqing (lower right). The histograms are based on the city-level daily API scores during 2001-2010. Scores greater than 200 are not illustrated. The red lines denote the cut-off points for each pollution level. Levels I and II are defined as "blue-sky days."

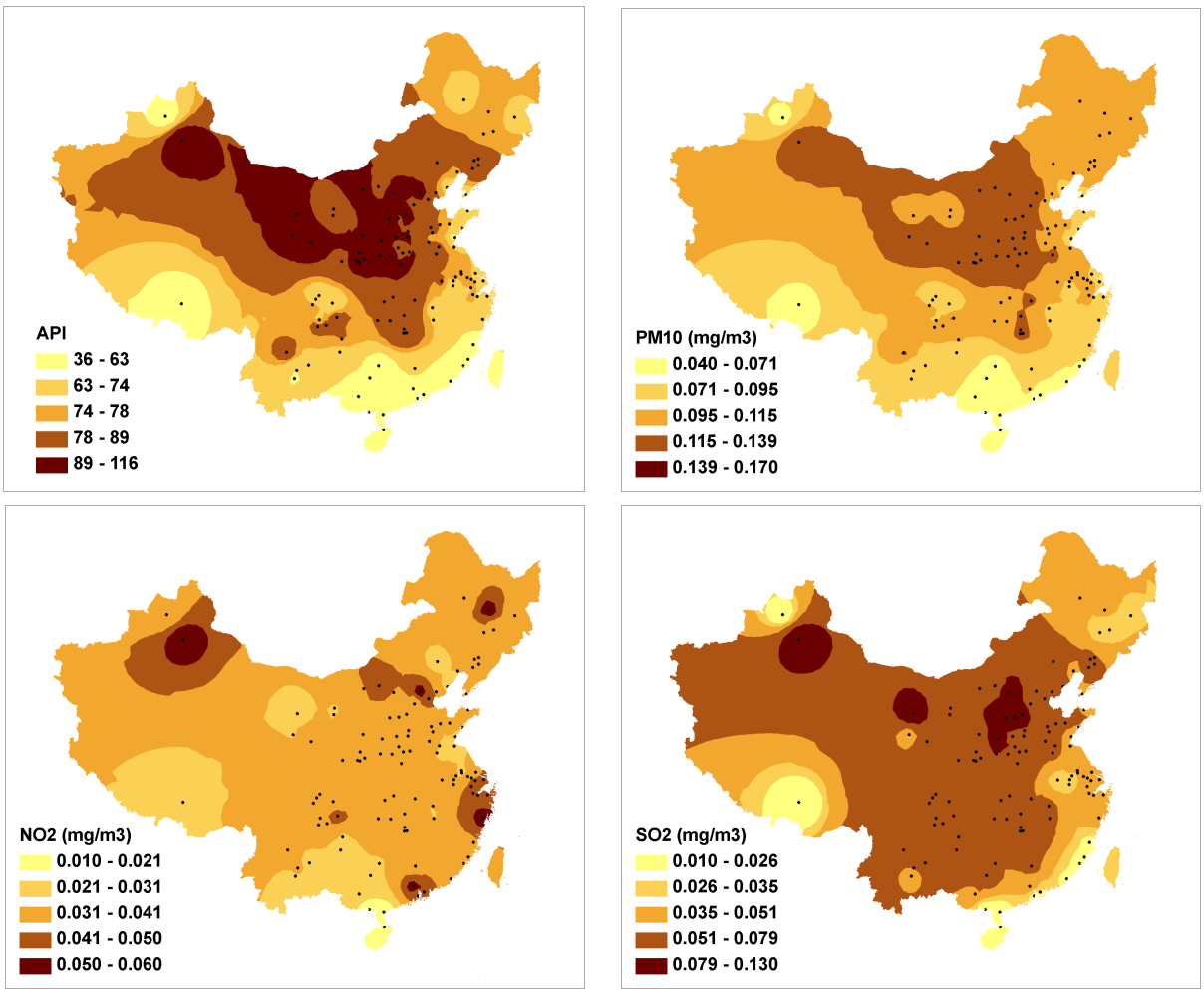


Figure 2: Average daily air pollution levels during 2001-2010: air pollution index (API, upper left), particulate matter concentration (PM_{10}, upper right), nitrogen dioxide concentration (NO_2, lower left), and sulfur dioxide concentration (SO_2, lower right). The filled contour plot shows spatially interpolated air pollution levels, which is generated by inverse distance weighting approach based on the city-level daily air pollution data. The dots represent cities that disclose daily API scores.

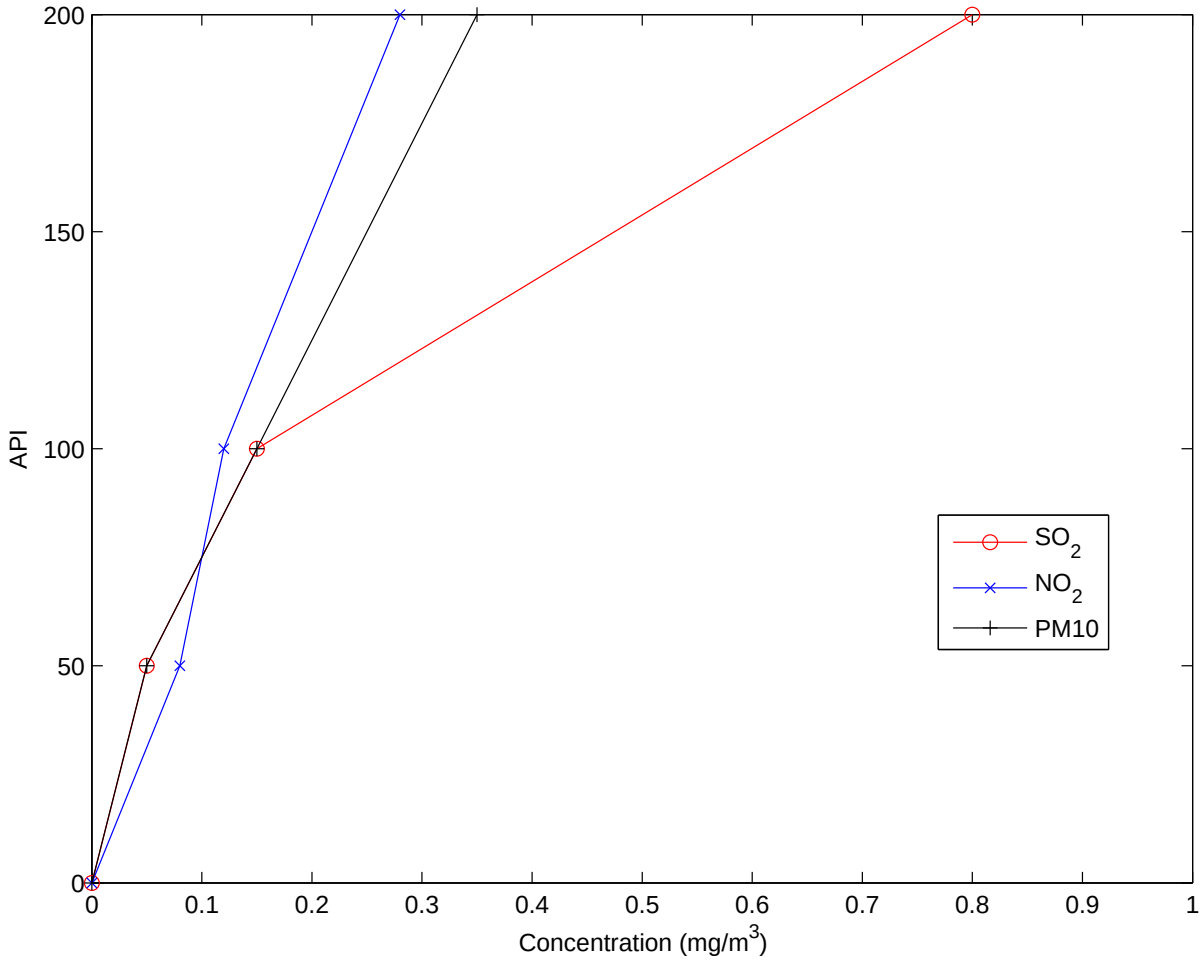


Figure 3: API is a nonlinear transformation of pollutant concentrations. In particular, there are kinks corresponding to API=100 for SO_2 and NO_2 respectively

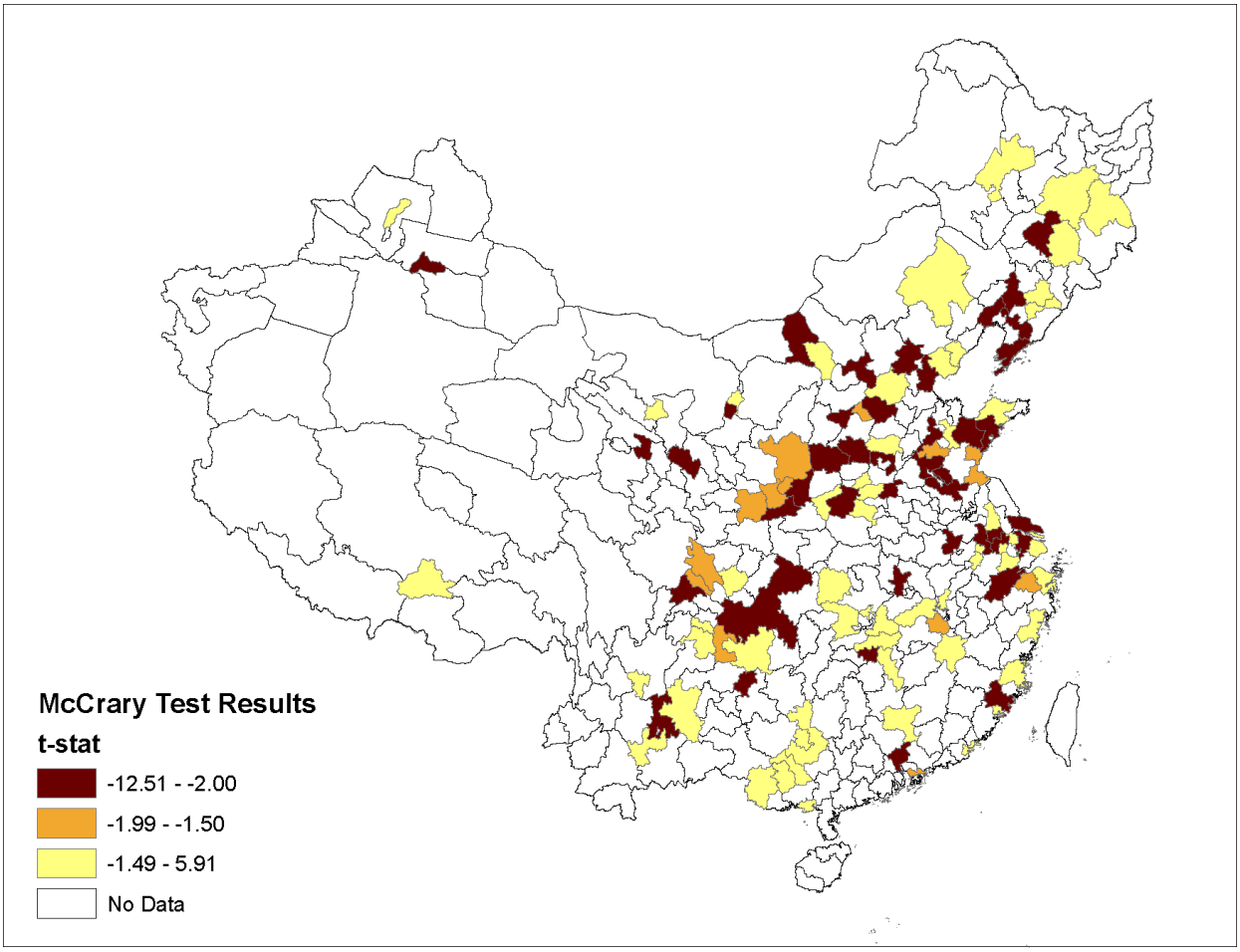


Figure 4: Results of McCrary (2008) tests show different degrees of manipulation in the PM_{10} data. Three levels of manipulation are illustrated: t-stats below -1.5 as non-manipulators, t-stats between -1.5 and -2 as borderline manipulators, and finally, t-stats smaller than -2 as manipulators. PM_{10} is selected because it is the dominant primary pollutant. A polygon is a prefecture that has the city as its capital.

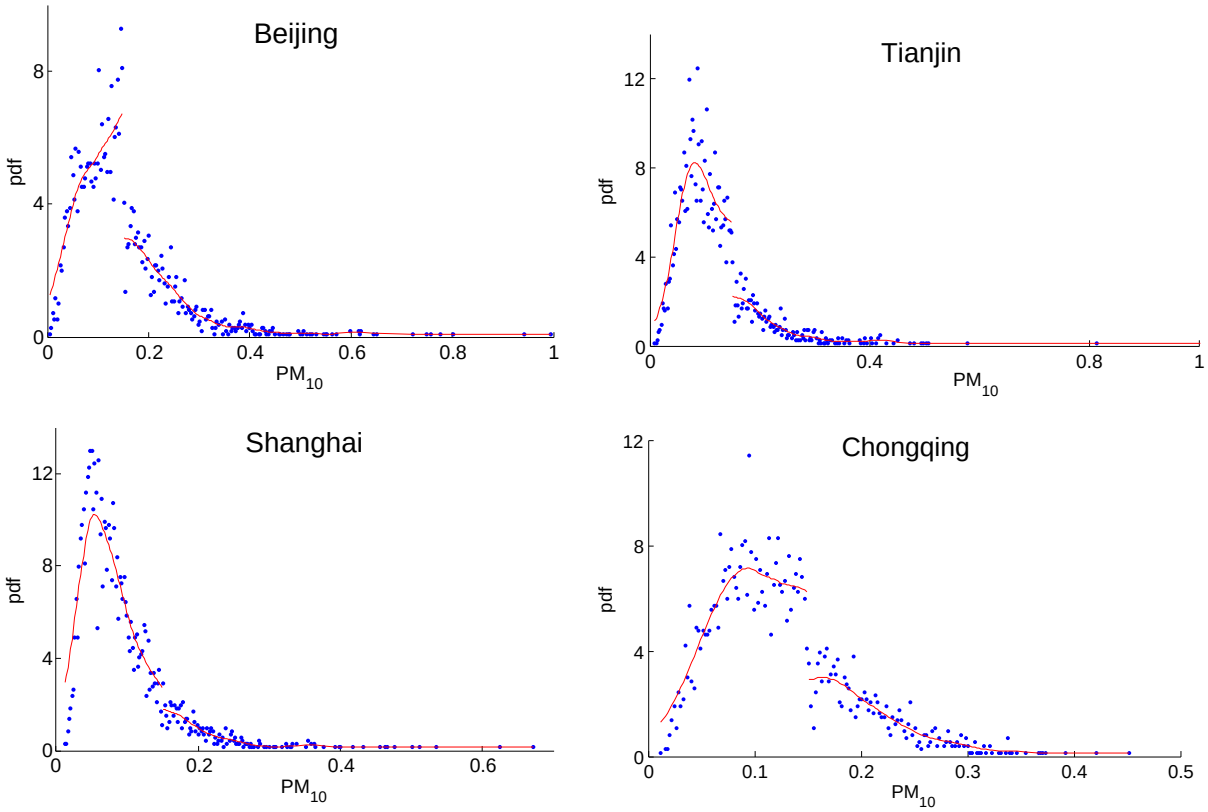


Figure 5: The McCrary (2008) test on four major cities. Beijing (upper left), Tianjin (upper right), Shanghai (lower left), and Chongqing (lower right). The blue dots are the first-step histogram obtained from implementing McCrary (2008) on the city-level daily PM_{10} concentrations during 2001-2010. The red curve is the Kernel estimator of the pdf to the right and left of the cut-off.

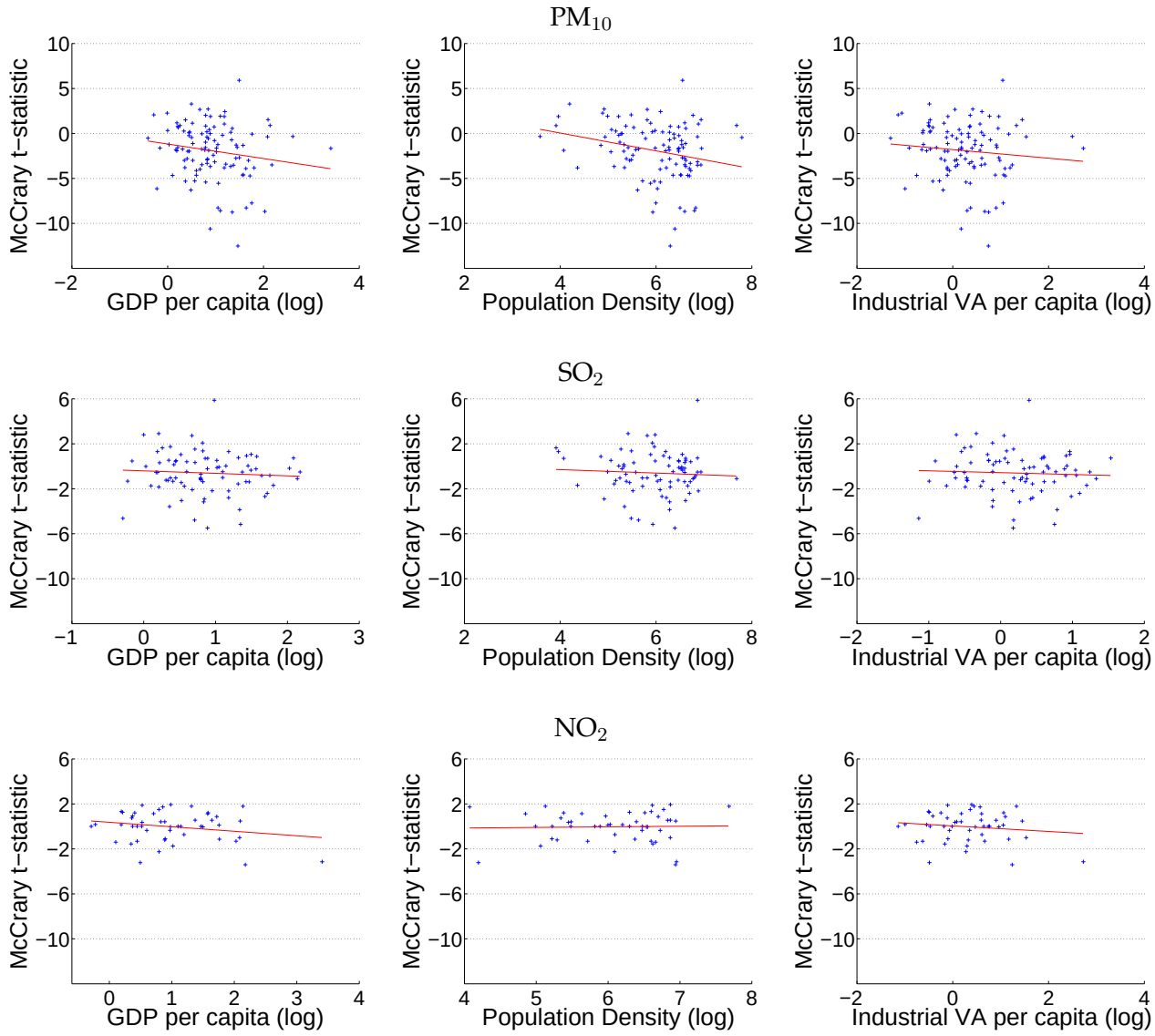


Figure 6: GDP per capita, population density, industrial value added per capita and McCrary results. The scatter plots graph the relationship between the log of GDP per capita (column 1), log of population density (column 2), log of value added of the secondary industry per capita (column 3) and the McCrary t-statistics for PM_{10} (upper), SO_2 (middle), and NO_2 (lower).

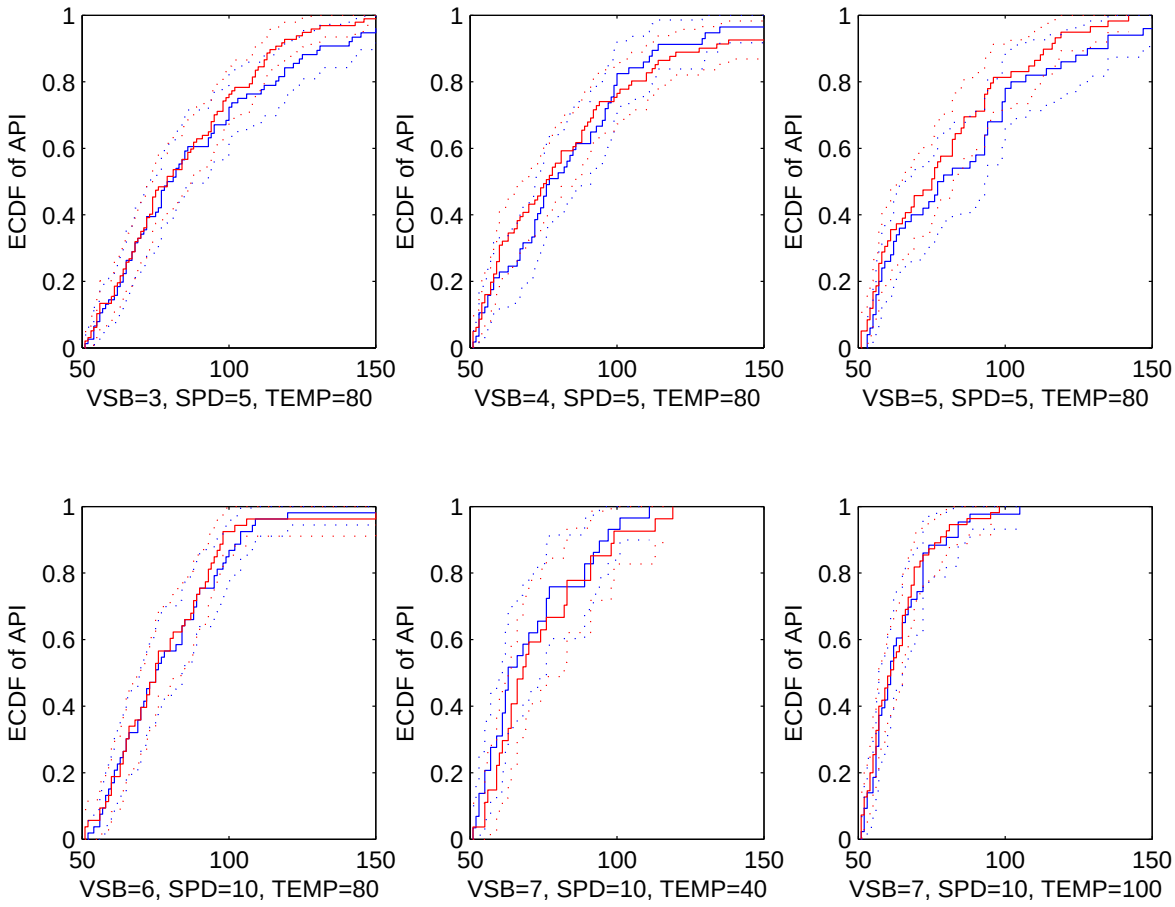


Figure 7: **Panel Matching Approach: Zhenjiang-Yangzhou.** The red lines represent Zhenjiang, the blue Yangzhou. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

'Effortless Perfection:'
Do Chinese Cities Manipulate Air Pollution Data?

Supplementary Appendix

April 25, 2013

Table SA.1: Kaifeng-Zhengzhou I

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	40	0	67	71	0.0625	0.0193	0.0208	DNR	DNR	DNR
1	5	60	0	27	28	0.4027	0.4698	0.4730	DNR	DNR	DNR
1	5	80	0	36	36	0.6585	0.3389	0.3422	DNR	DNR	DNR
1	10	40	0	32	32	0.9509	0.8767	0.8772	DNR	DNR	DNR
1	10	60	0	11	11	0.7358	0.9653	0.9657	DNR	DNR	DNR
2	5	40	0	60	51	0.0228	0.0147	0.0163	DNR	DNR	DNR
2	5	60	0	73	76	0.1498	0.0196	0.0210	DNR	DNR	DNR
2	5	80	0	97	99	0.9859	0.3739	0.3750	DNR	DNR	DNR
2	5	100	0	26	26	0.1383	0.0057	0.0079	DNR	DNR	DNR
2	10	40	0	23	20	0.6491	0.4628	0.4669	DNR	DNR	DNR
2	10	60	0	31	32	0.9078	0.3176	0.3215	DNR	DNR	DNR
2	10	80	0	36	38	0.8405	0.9974	0.9974	DNR	DNR	DNR
3	5	40	0	53	41	0.0753	0.0040	0.0050	DNR	DNR	DNR
3	5	60	0	87	93	0.0934	0.0365	0.0379	DNR	DNR	DNR
3	5	80	0	150	163	0.7569	0.1216	0.1227	DNR	DNR	DNR
3	5	100	0	40	44	0.3214	0.3613	0.3640	DNR	DNR	DNR
3	10	40	0	16	17	0.9469	0.5879	0.5918	DNR	DNR	DNR
3	10	60	0	59	62	0.5417	0.7185	0.7192	DNR	DNR	DNR
3	10	80	0	56	63	0.9860	0.9705	0.9705	DNR	DNR	DNR
3	10	100	0	16	17	0.1049	0.3295	0.3371	DNR	DNR	DNR
4	5	40	0	35	39	0.0080	0.0014	0.0020	DNR	DNR	DNR
4	5	60	0	51	53	0.0157	0.0011	0.0015	DNR	DNR	DNR
4	5	80	0	124	147	0.0064	0.0117	0.0123	DNR	DNR	DNR
4	5	100	0	26	35	0.0339	0.0013	0.0020	DNR	DNR	DNR
4	10	40	0	20	20	0.4973	0.9042	0.9049	DNR	DNR	DNR
4	10	60	0	66	64	0.9637	0.7458	0.7463	DNR	DNR	DNR
4	10	80	0	79	87	0.7162	0.6505	0.6511	DNR	DNR	DNR
4	10	100	0	20	23	0.7116	0.8739	0.8747	DNR	DNR	DNR
4	15	60	0	11	11	0.3744	0.7812	0.7841	DNR	DNR	DNR
5	5	40	0	31	33	0.0491	0.0127	0.0154	DNR	DNR	DNR
5	5	60	0	43	51	0.1422	0.0245	0.0269	DNR	DNR	DNR
5	5	80	0	86	125	0.0000	0.0022	0.0025	R	DNR	DNR
5	5	100	0	39	44	0.0639	0.0075	0.0091	DNR	DNR	DNR
5	10	40	0	25	20	0.0717	0.1633	0.1705	DNR	DNR	DNR
5	10	60	0	51	51	0.6902	0.2137	0.2166	DNR	DNR	DNR
5	10	80	0	40	51	0.0876	0.0420	0.0450	DNR	DNR	DNR
5	10	100	0	27	30	0.3178	0.5749	0.5771	DNR	DNR	DNR
5	15	60	0	11	13	0.9427	0.7800	0.7826	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.2: Kaifeng-Zhengzhou II

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
6	5	40	0	35	43	0.0112	0.0101	0.0120	DNR	DNR	DNR
6	5	60	0	51	86	0.0003	0.0004	0.0005	R	R	DNR
6	5	80	0	93	133	0.0000	0.0000	0.0000	R	R	R
6	5	100	0	51	65	0.0017	0.0002	0.0004	DNR	R	R
6	10	40	0	16	19	0.7035	0.3203	0.3275	DNR	DNR	DNR
6	10	60	0	48	60	0.8959	0.8536	0.8539	DNR	DNR	DNR
6	10	80	0	60	77	0.5433	0.1698	0.1721	DNR	DNR	DNR
6	10	100	0	19	23	0.6399	0.6476	0.6501	DNR	DNR	DNR
7	5	40	0	33	45	0.0009	0.0001	0.0002	DNR	R	R
7	5	60	0	55	83	0.0000	0.0002	0.0003	R	R	R
7	5	80	0	75	111	0.0000	0.0000	0.0000	R	R	R
7	5	100	0	40	66	0.0001	0.0000	0.0000	R	R	R
7	10	40	0	37	45	0.0000	0.0001	0.0001	R	R	R
7	10	60	0	60	89	0.0035	0.0002	0.0003	DNR	R	R
7	10	80	0	79	98	0.0034	0.0049	0.0055	DNR	DNR	DNR
7	10	100	0	32	37	0.1033	0.0098	0.0120	DNR	DNR	DNR
7	15	40	0	14	12	0.9533	0.4093	0.4174	DNR	DNR	DNR
7	15	60	0	11	14	0.7322	0.4488	0.4565	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.3: Zhuzhou-Xiangtan

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	40	0	28	26	0.7573	0.7322	0.7335	DNR	DNR	DNR
1	5	60	0	50	51	0.4248	0.4811	0.4828	DNR	DNR	DNR
1	5	80	0	15	16	0.7925	0.6536	0.6570	DNR	DNR	DNR
1	10	40	0	17	22	0.0228	0.0871	0.0954	DNR	DNR	DNR
1	10	60	0	49	50	0.4558	0.1964	0.1994	DNR	DNR	DNR
1	10	80	0	18	21	0.6233	0.9315	0.9320	DNR	DNR	DNR
2	5	60	0	71	65	0.0230	0.1160	0.1184	DNR	DNR	DNR
2	5	80	0	48	48	0.4803	0.3849	0.3871	DNR	DNR	DNR
2	10	40	0	24	22	0.2500	0.5233	0.5266	DNR	DNR	DNR
2	10	60	0	60	60	0.4766	0.5376	0.5388	DNR	DNR	DNR
2	10	80	0	40	43	0.1353	0.0932	0.0971	DNR	DNR	DNR
3	5	60	0	39	38	0.0118	0.2559	0.2596	DNR	DNR	DNR
3	5	80	0	44	52	0.2878	0.0476	0.0505	DNR	DNR	DNR
3	10	60	0	26	27	0.3274	0.9595	0.9597	DNR	DNR	DNR
3	10	80	0	40	42	0.9999	0.6264	0.6277	DNR	DNR	DNR
4	5	60	0	50	47	0.0350	0.4529	0.4547	DNR	DNR	DNR
4	5	80	0	37	42	0.1975	0.1394	0.1435	DNR	DNR	DNR
4	5	100	0	14	15	0.2118	0.1114	0.1230	DNR	DNR	DNR
4	10	60	0	24	21	0.1315	0.4035	0.4081	DNR	DNR	DNR
4	10	80	0	22	27	0.8399	0.4443	0.4481	DNR	DNR	DNR
5	5	60	0	44	46	0.1810	0.6914	0.6924	DNR	DNR	DNR
5	5	80	0	79	78	0.0073	0.0248	0.0262	DNR	DNR	DNR
5	5	100	0	21	18	0.2577	0.0704	0.0785	DNR	DNR	DNR
5	10	80	0	29	32	0.9683	0.9895	0.9895	DNR	DNR	DNR
6	5	60	0	37	38	0.0124	0.0072	0.0089	DNR	DNR	DNR
6	5	80	0	79	75	0.2991	0.2315	0.2334	DNR	DNR	DNR
6	5	100	0	32	38	0.5879	0.0914	0.0960	DNR	DNR	DNR
6	10	80	0	31	31	0.9439	0.6632	0.6648	DNR	DNR	DNR
6	10	100	0	19	19	0.9563	0.9112	0.9118	DNR	DNR	DNR
7	5	40	0	19	21	0.8778	0.7544	0.7561	DNR	DNR	DNR
7	5	60	0	183	173	0.0000	0.0005	0.0006	R	R	R
7	5	80	0	256	273	0.0000	0.0001	0.0001	R	R	R
7	5	100	0	242	208	0.0496	0.0232	0.0236	DNR	DNR	DNR
7	10	40	0	23	19	0.2185	0.0501	0.0571	DNR	DNR	DNR
7	10	60	0	64	56	0.0256	0.3158	0.3178	DNR	DNR	DNR
7	10	80	0	100	98	0.1975	0.2279	0.2294	DNR	DNR	DNR
7	10	100	0	140	96	0.0066	0.0018	0.0020	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.4: Wuhu-Maanshan

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
2	5	60	0	13	106	0.5487	0.8655	0.8658	DNR	DNR	DNR
3	5	60	0	16	65	0.7582	0.4673	0.4695	DNR	DNR	DNR
3	5	80	0	11	84	0.9126	0.2646	0.2675	DNR	DNR	DNR
4	5	40	0	13	20	0.7181	0.5097	0.5145	DNR	DNR	DNR
4	5	60	0	41	53	0.1476	0.6236	0.6248	DNR	DNR	DNR
4	5	80	0	20	66	0.9683	0.6243	0.6256	DNR	DNR	DNR
4	10	60	0	20	52	0.1043	0.0034	0.0046	DNR	DNR	DNR
4	10	80	0	13	94	0.8048	0.5232	0.5246	DNR	DNR	DNR
5	5	40	0	22	19	0.1159	0.5460	0.5495	DNR	DNR	DNR
5	5	60	0	57	29	0.7715	0.6866	0.6876	DNR	DNR	DNR
5	5	80	0	61	58	0.9391	0.9543	0.9544	DNR	DNR	DNR
5	10	60	0	33	34	0.9092	0.7046	0.7058	DNR	DNR	DNR
5	10	80	0	24	82	0.1363	0.9957	0.9957	DNR	DNR	DNR
6	5	40	0	18	27	0.8118	0.6364	0.6388	DNR	DNR	DNR
6	5	60	0	84	24	0.0377	0.1528	0.1558	DNR	DNR	DNR
6	5	80	0	104	32	0.9095	0.4026	0.4041	DNR	DNR	DNR
6	5	100	0	35	13	0.4935	0.2005	0.2070	DNR	DNR	DNR
6	10	40	0	17	15	0.7471	0.9640	0.9643	DNR	DNR	DNR
6	10	60	0	57	41	0.7058	0.8936	0.8939	DNR	DNR	DNR
6	10	80	0	78	52	0.8807	0.9414	0.9415	DNR	DNR	DNR
6	10	100	0	22	28	0.7090	0.0702	0.0765	DNR	DNR	DNR
7	5	40	0	73	15	0.4402	0.8398	0.8402	DNR	DNR	DNR
7	10	40	0	50	34	0.6852	0.3332	0.3360	DNR	DNR	DNR
7	10	60	0	119	25	0.6665	0.4891	0.4902	DNR	DNR	DNR
7	10	80	0	167	29	0.5148	0.9954	0.9954	DNR	DNR	DNR
7	10	100	0	56	31	0.2858	0.9412	0.9414	DNR	DNR	DNR
7	15	100	0	15	12	0.0752	0.0666	0.0785	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.5: Quanzhou-Xiamen

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	80	0	12	12	0.9913	0.5232	0.5298	DNR	DNR	DNR
2	5	80	0	27	30	0.9477	0.8047	0.8056	DNR	DNR	DNR
3	5	80	0	67	70	0.5121	0.5612	0.5622	DNR	DNR	DNR
3	10	80	0	20	20	0.4973	0.7779	0.7794	DNR	DNR	DNR
4	5	80	0	87	108	0.3514	0.0896	0.0913	DNR	DNR	DNR
4	5	100	0	14	15	0.0022	0.0019	0.0044	DNR	DNR	DNR
4	10	60	0	12	15	0.3725	0.7690	0.7714	DNR	DNR	DNR
4	10	80	0	44	45	0.1632	0.7548	0.7555	DNR	DNR	DNR
5	5	60	0	15	22	0.8380	0.5844	0.5879	DNR	DNR	DNR
5	5	80	0	97	125	0.2688	0.0432	0.0444	DNR	DNR	DNR
5	5	100	0	30	38	0.0120	0.0003	0.0005	DNR	R	R
5	10	60	0	33	36	0.3630	0.8608	0.8613	DNR	DNR	DNR
5	10	80	0	76	93	0.4593	0.0655	0.0673	DNR	DNR	DNR
5	15	80	0	12	12	0.4333	0.2306	0.2433	DNR	DNR	DNR
6	5	60	0	25	35	0.0892	0.1698	0.1751	DNR	DNR	DNR
6	5	80	0	77	114	0.9568	0.7187	0.7191	DNR	DNR	DNR
6	5	100	0	52	56	0.0037	0.0003	0.0005	DNR	R	R
6	10	60	0	40	60	0.7569	0.7920	0.7925	DNR	DNR	DNR
6	10	80	0	116	154	0.0922	0.0932	0.0943	DNR	DNR	DNR
6	10	100	0	46	41	0.0037	0.0001	0.0001	DNR	R	R
6	15	60	0	17	22	0.2700	0.2260	0.2337	DNR	DNR	DNR
6	15	80	0	19	25	0.3033	0.2033	0.2103	DNR	DNR	DNR
7	5	60	0	12	22	0.7611	0.5816	0.5854	DNR	DNR	DNR
7	5	80	0	50	65	0.0025	0.0314	0.0335	DNR	DNR	DNR
7	5	100	0	94	48	0.0001	0.0000	0.0000	R	R	R
7	10	60	0	73	109	0.1008	0.3148	0.3161	DNR	DNR	DNR
7	10	80	0	181	307	0.2900	0.7290	0.7291	DNR	DNR	DNR
7	10	100	0	292	132	0.0000	0.0000	0.0000	R	R	R
7	15	60	0	31	48	0.7224	0.1639	0.1679	DNR	DNR	DNR
7	15	80	0	64	91	0.1804	0.4771	0.4781	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.6: Zhenjiang-Yangzhou

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	40	0	14	23	0.8120	0.3968	0.4025	DNR	DNR	DNR
1	5	60	0	71	85	0.6679	0.5075	0.5085	DNR	DNR	DNR
1	5	80	0	50	53	0.9781	0.7284	0.7291	DNR	DNR	DNR
1	10	60	0	35	44	0.8856	0.5603	0.5620	DNR	DNR	DNR
1	10	80	0	13	16	0.8811	0.4607	0.4671	DNR	DNR	DNR
2	5	40	0	28	43	0.2455	0.1794	0.1838	DNR	DNR	DNR
2	5	60	0	98	122	0.5607	0.4785	0.4793	DNR	DNR	DNR
2	5	80	0	89	104	0.7897	0.9044	0.9045	DNR	DNR	DNR
2	5	100	0	24	29	0.8157	0.6734	0.6752	DNR	DNR	DNR
2	10	40	0	19	27	0.7318	0.7817	0.7830	DNR	DNR	DNR
2	10	60	0	68	86	0.9782	0.8893	0.8895	DNR	DNR	DNR
2	10	80	0	59	80	0.7626	0.6522	0.6529	DNR	DNR	DNR
2	10	100	0	27	30	0.9680	0.8359	0.8366	DNR	DNR	DNR
3	5	40	0	19	27	0.5722	0.5288	0.5320	DNR	DNR	DNR
3	5	60	0	65	78	0.7321	0.5281	0.5292	DNR	DNR	DNR
3	5	80	0	76	97	0.6844	0.2391	0.2408	DNR	DNR	DNR
3	5	100	0	57	70	0.9912	0.8622	0.8625	DNR	DNR	DNR
3	10	40	0	16	23	0.8069	0.5568	0.5604	DNR	DNR	DNR
3	10	60	0	49	63	0.3699	0.4071	0.4089	DNR	DNR	DNR
3	10	80	0	74	91	0.7134	0.8436	0.8438	DNR	DNR	DNR
3	10	100	0	25	30	0.8441	0.6971	0.6986	DNR	DNR	DNR
4	5	40	0	14	22	0.1349	0.0914	0.1005	DNR	DNR	DNR
4	5	60	0	48	56	0.7372	0.3547	0.3569	DNR	DNR	DNR
4	5	80	0	57	81	0.6976	0.7861	0.7866	DNR	DNR	DNR
4	5	100	0	27	27	0.4656	0.8790	0.8796	DNR	DNR	DNR
4	10	60	0	51	53	0.9203	0.9195	0.9197	DNR	DNR	DNR
4	10	80	0	81	99	0.5775	0.2602	0.2618	DNR	DNR	DNR
4	10	100	0	27	35	0.9859	0.6564	0.6580	DNR	DNR	DNR
5	5	40	0	14	18	0.2120	0.1927	0.2026	DNR	DNR	DNR
5	5	60	0	27	42	0.6549	0.9746	0.9747	DNR	DNR	DNR
5	5	80	0	50	59	0.4997	0.1796	0.1824	DNR	DNR	DNR
5	5	100	0	22	22	0.9786	0.6450	0.6473	DNR	DNR	DNR
5	10	60	0	31	40	0.7884	0.2329	0.2369	DNR	DNR	DNR
5	10	80	0	79	96	0.9690	0.5792	0.5799	DNR	DNR	DNR
5	10	100	0	33	40	0.7976	0.3886	0.3915	DNR	DNR	DNR
6	5	40	0	17	27	0.9121	0.8383	0.8393	DNR	DNR	DNR
6	5	60	0	25	26	0.7969	0.9496	0.9499	DNR	DNR	DNR
6	5	80	0	25	25	0.6485	0.3830	0.3874	DNR	DNR	DNR
6	5	100	0	19	20	0.1748	0.0412	0.0484	DNR	DNR	DNR
6	10	60	0	39	45	0.7092	0.5945	0.5959	DNR	DNR	DNR
6	10	80	0	53	53	0.9646	0.7828	0.7834	DNR	DNR	DNR
6	10	100	0	33	39	0.4404	0.3402	0.3435	DNR	DNR	DNR
7	5	100	0	11	11	0.9852	0.7312	0.7348	DNR	DNR	DNR
7	10	40	0	29	27	0.6859	0.4594	0.4626	DNR	DNR	DNR
7	10	60	0	21	22	0.9156	0.9328	0.9332	DNR	DNR	DNR
7	10	80	0	23	24	0.9400	0.3782	0.3829	DNR	DNR	DNR
7	10	100	0	43	55	0.9691	0.6984	0.6993	DNR	DNR	DNR
7	15	100	0	14	12	0.8900	0.9693	0.9696	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.7: Changzhou-Wuxi

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	60	0	14	13	0.6139	0.4343	0.4416	DNR	DNR	DNR
2	5	60	0	25	22	0.9975	0.4736	0.4773	DNR	DNR	DNR
3	5	40	0	16	12	0.7193	0.3112	0.3205	DNR	DNR	DNR
3	5	60	0	40	34	0.9252	0.6589	0.6602	DNR	DNR	DNR
3	5	80	0	34	33	0.9790	0.8987	0.8991	DNR	DNR	DNR
3	10	60	0	24	24	0.6216	0.4190	0.4232	DNR	DNR	DNR
4	5	40	0	19	13	0.5109	0.1336	0.1440	DNR	DNR	DNR
4	5	60	0	66	55	0.9280	0.7815	0.7820	DNR	DNR	DNR
4	5	80	0	39	37	0.2888	0.1790	0.1831	DNR	DNR	DNR
4	10	60	0	21	24	0.8857	0.7477	0.7492	DNR	DNR	DNR
4	10	60	0.5	11	11	0.7358	0.7414	0.7448	DNR	DNR	DNR
4	10	80	0	14	13	0.9410	0.6197	0.6240	DNR	DNR	DNR
5	5	40	0	21	17	0.8064	0.1417	0.1504	DNR	DNR	DNR
5	5	60	0	78	74	0.2797	0.4316	0.4329	DNR	DNR	DNR
5	5	80	0	67	69	0.2734	0.1035	0.1059	DNR	DNR	DNR
5	5	100	0	23	24	0.9962	0.8966	0.8971	DNR	DNR	DNR
5	10	60	0	26	24	0.8110	0.8012	0.8023	DNR	DNR	DNR
5	10	80	0	26	23	0.7694	0.9638	0.9640	DNR	DNR	DNR
6	5	40	0	30	22	0.5276	0.6185	0.6207	DNR	DNR	DNR
6	5	60	0	110	95	0.8838	0.6282	0.6287	DNR	DNR	DNR
6	5	80	0	118	113	0.0965	0.0630	0.0643	DNR	DNR	DNR
6	5	100	0	50	50	0.0951	0.1180	0.1212	DNR	DNR	DNR
6	10	60	0	36	29	0.9887	0.8685	0.8690	DNR	DNR	DNR
6	10	80	0	39	38	0.3007	0.0768	0.0809	DNR	DNR	DNR
7	5	40	0	147	113	0.9164	0.4551	0.4558	DNR	DNR	DNR
7	5	60	0	257	233	0.1258	0.4931	0.4934	DNR	DNR	DNR
7	5	80	0	444	412	0.0763	0.0455	0.0459	DNR	DNR	DNR
7	5	100	0	295	258	0.0423	0.0094	0.0097	DNR	DNR	DNR
7	5	100	0.5	14	13	0.8124	0.6101	0.6146	DNR	DNR	DNR
7	10	40	0	56	45	0.3585	0.4092	0.4112	DNR	DNR	DNR
7	10	60	0	126	119	0.9074	0.8903	0.8905	DNR	DNR	DNR
7	10	80	0	176	168	0.1355	0.0843	0.0852	DNR	DNR	DNR
7	10	100	0	116	97	0.1525	0.0896	0.0911	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.8: Jinan-Taian

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	40	0	41	33	0.0000	0.0000	0.0000	R	R	R
1	5	60	0	17	12	0.0000	0.0000	0.0002	R	R	R
2	5	40	0	17	14	0.0002	0.0000	0.0000	R	R	R
2	5	80	0	28	18	0.0458	0.0091	0.0124	DNR	DNR	DNR
3	5	40	0	21	21	0.0003	0.0000	0.0002	R	R	R
3	5	80	0	39	25	0.0858	0.0396	0.0438	DNR	DNR	DNR
3	5	100	0	14	13	0.0051	0.1064	0.1190	DNR	DNR	DNR
3	10	40	0	15	11	0.0678	0.0050	0.0097	DNR	DNR	DNR
3	10	80	0	24	17	0.1816	0.0216	0.0271	DNR	DNR	DNR
4	5	60	0	21	15	0.0097	0.0030	0.0054	DNR	DNR	DNR
4	5	80	0	36	29	0.0000	0.0000	0.0000	R	R	R
4	10	40	0	23	14	0.0377	0.0099	0.0142	DNR	DNR	DNR
4	10	60	0	20	14	0.0090	0.0004	0.0012	DNR	R	R
4	10	80	0	37	31	0.0032	0.0003	0.0005	DNR	R	R
5	5	80	0	44	31	0.0002	0.0000	0.0000	R	R	R
5	10	40	0	26	22	0.0061	0.0013	0.0024	DNR	R	DNR
5	10	60	0	29	18	0.0000	0.0117	0.0154	R	DNR	DNR
5	10	80	0	44	24	0.0007	0.0006	0.0010	R	R	R
5	10	100	0	32	27	0.0002	0.0001	0.0003	R	R	R
6	5	40	0	16	13	0.1952	0.1235	0.1352	DNR	DNR	DNR
6	5	80	0	48	34	0.0078	0.0029	0.0039	DNR	DNR	DNR
6	5	100	0	13	12	0.0145	0.0096	0.0164	DNR	DNR	DNR
6	10	40	0	26	17	0.0116	0.0105	0.0142	DNR	DNR	DNR
6	10	60	0	34	19	0.0006	0.0162	0.0199	R	DNR	DNR
6	10	80	0	65	45	0.0000	0.0000	0.0000	R	R	R
6	10	100	0	29	27	0.0000	0.0000	0.0001	R	R	R
7	5	40	0	83	82	0.0000	0.0000	0.0000	R	R	R
7	5	60	0	86	55	0.0111	0.0014	0.0018	DNR	R	DNR
7	5	80	0	138	94	0.0000	0.0000	0.0000	R	R	R
7	5	100	0	40	28	0.0001	0.0001	0.0003	R	R	R
7	10	40	0	252	204	0.0000	0.0000	0.0000	R	R	R
7	10	60	0	433	310	0.0000	0.0000	0.0000	R	R	R
7	10	80	0	518	388	0.0000	0.0000	0.0000	R	R	R
7	10	100	0	193	147	0.0000	0.0000	0.0000	R	R	R
7	15	40	0	28	18	0.7769	0.4379	0.4421	DNR	DNR	DNR
7	15	60	0	93	67	0.0000	0.0002	0.0002	R	R	R
7	15	80	0	136	106	0.0000	0.0000	0.0000	R	R	R
7	15	100	0	59	47	0.0012	0.0133	0.0149	R	DNR	DNR
7	20	80	0	21	18	0.0048	0.0008	0.0019	DNR	R	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.9: Hangzhou-Shaoxing

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	40	0	12	11	0.9727	0.7364	0.7398	DNR	DNR	DNR
1	5	60	0	141	111	0.1059	0.0466	0.0477	DNR	DNR	DNR
1	5	80	0	58	49	0.5123	0.1239	0.1269	DNR	DNR	DNR
1	10	60	0	28	18	0.1781	0.0907	0.0977	DNR	DNR	DNR
1	10	80	0	14	11	0.5044	0.9622	0.9626	DNR	DNR	DNR
2	5	40	0	36	30	0.4830	0.6945	0.6958	DNR	DNR	DNR
2	5	60	0	196	146	0.0091	0.0553	0.0561	DNR	DNR	DNR
2	5	80	0	157	124	0.0192	0.0094	0.0099	DNR	DNR	DNR
2	5	100	0	35	31	0.3738	0.0407	0.0448	DNR	DNR	DNR
2	10	40	0	13	14	0.7158	0.3001	0.3100	DNR	DNR	DNR
2	10	60	0	72	57	0.6735	0.9000	0.9002	DNR	DNR	DNR
2	10	60	0.5	19	12	0.6172	0.3176	0.3259	DNR	DNR	DNR
2	10	80	0	46	37	0.8626	0.9949	0.9950	DNR	DNR	DNR
3	5	40	0	19	14	0.9904	0.7114	0.7139	DNR	DNR	DNR
3	5	60	0	106	86	0.0237	0.0214	0.0225	DNR	DNR	DNR
3	5	80	0	144	107	0.0046	0.0030	0.0033	DNR	DNR	DNR
3	5	100	0	59	52	0.0089	0.0392	0.0415	DNR	DNR	DNR
3	10	60	0	58	46	0.4941	0.8655	0.8658	DNR	DNR	DNR
3	10	80	0	56	43	0.0015	0.0011	0.0016	DNR	DNR	DNR
3	10	100	0	15	14	0.2118	0.2278	0.2383	DNR	DNR	DNR
4	5	40	0	20	12	0.8251	0.8652	0.8663	DNR	DNR	DNR
4	5	60	0	89	66	0.2322	0.1983	0.2002	DNR	DNR	DNR
4	5	80	0	136	101	0.0039	0.0003	0.0004	DNR	R	R
4	5	100	0	68	53	0.0015	0.0011	0.0014	DNR	DNR	DNR
4	10	60	0	56	37	0.2934	0.0981	0.1016	DNR	DNR	DNR
4	10	80	0	93	73	0.1606	0.0360	0.0375	DNR	DNR	DNR
4	10	100	0	15	12	0.1462	0.0602	0.0719	DNR	DNR	DNR
5	5	40	0	24	16	0.6344	0.1257	0.1340	DNR	DNR	DNR
5	5	60	0	59	42	0.0022	0.0053	0.0064	DNR	DNR	DNR
5	5	80	0	128	87	0.0001	0.0008	0.0009	R	DNR	DNR
5	5	100	0	55	33	0.0033	0.0001	0.0002	DNR	R	R
5	10	60	0	50	38	0.5495	0.7592	0.7600	DNR	DNR	DNR
5	10	80	0	90	62	0.0587	0.0225	0.0239	DNR	DNR	DNR
5	10	100	0	27	20	0.8768	0.3209	0.3262	DNR	DNR	DNR
6	5	40	0	18	12	0.9782	0.8296	0.8311	DNR	DNR	DNR
6	5	60	0	41	29	0.9668	0.9181	0.9184	DNR	DNR	DNR
6	5	80	0	56	31	0.0006	0.0160	0.0181	DNR	DNR	DNR
6	5	100	0	74	43	0.0000	0.0000	0.0000	R	R	R
6	10	40	0	23	18	0.9933	0.7897	0.7911	DNR	DNR	DNR
6	10	60	0	47	39	0.3332	0.3335	0.3362	DNR	DNR	DNR
6	10	80	0	98	67	0.1891	0.1249	0.1269	DNR	DNR	DNR
6	10	100	0	56	38	0.0010	0.0001	0.0002	DNR	R	R
7	10	40	0	20	17	0.3430	0.8766	0.8775	DNR	DNR	DNR
7	10	60	0	21	15	0.9379	0.9908	0.9908	DNR	DNR	DNR
7	10	80	0	37	20	0.3750	0.1791	0.1846	DNR	DNR	DNR
7	10	100	0	96	30	0.2236	0.1714	0.1739	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.10: Huhehaote-Baotou

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
7	5	20	0	85	174	0.0000	0.0000	0.0000	R	R	R
7	5	40	0	165	260	0.0000	0.0000	0.0000	R	R	R
7	5	60	0	209	238	0.0000	0.0000	0.0000	R	R	R
7	5	80	0	391	314	0.0000	0.0000	0.0000	R	R	R
7	10	20	0	87	189	0.0029	0.0011	0.0012	R	R	R
7	10	40	0	167	404	0.0000	0.0000	0.0000	R	R	R
7	10	60	0	177	421	0.0000	0.0006	0.0006	R	R	R
7	10	80	0	237	605	0.0000	0.0000	0.0000	R	R	R
7	10	100	0	18	15	0.2633	0.8051	0.8067	DNR	DNR	DNR
7	15	20	0	15	18	0.2297	0.2742	0.2826	DNR	DNR	DNR
7	15	40	0	25	44	0.0004	0.0021	0.0030	R	R	R
7	15	60	0	21	49	0.5676	0.6243	0.6258	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.11: Jilin-Changchun

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
2	5	20	0	22	23	0.2876	0.0437	0.0500	DNR	DNR	DNR
3	5	20	0	55	55	0.2931	0.5001	0.5015	DNR	DNR	DNR
3	10	20	0	44	44	0.0019	0.0044	0.0055	DNR	DNR	DNR
3	10	40	0	22	21	0.5896	0.9046	0.9052	DNR	DNR	DNR
4	5	20	0	60	69	0.5941	0.6107	0.6116	DNR	DNR	DNR
4	5	40	0	14	16	0.9034	0.8758	0.8769	DNR	DNR	DNR
4	5	80	0	12	15	0.3725	0.8071	0.8091	DNR	DNR	DNR
4	10	20	0	78	80	0.1238	0.9665	0.9666	DNR	DNR	DNR
4	10	40	0	23	26	0.1220	0.2352	0.2411	DNR	DNR	DNR
4	10	60	0	14	14	0.8622	0.9148	0.9156	DNR	DNR	DNR
4	10	80	0	20	21	0.1090	0.2945	0.3010	DNR	DNR	DNR
5	5	20	0	27	37	0.6949	0.4855	0.4881	DNR	DNR	DNR
5	5	40	0	16	23	0.7231	0.4729	0.4774	DNR	DNR	DNR
5	5	80	0	23	28	0.0341	0.0223	0.0266	DNR	DNR	DNR
5	10	20	0	102	109	0.3150	0.9614	0.9615	DNR	DNR	DNR
5	10	40	0	35	42	0.2590	0.2268	0.2306	DNR	DNR	DNR
5	10	60	0	27	30	0.8707	0.7262	0.7275	DNR	DNR	DNR
5	10	80	0	41	48	0.5160	0.4272	0.4293	DNR	DNR	DNR
5	15	20	0	11	12	0.9094	0.4611	0.4693	DNR	DNR	DNR
5	15	40	0	18	20	0.0321	0.1370	0.1457	DNR	DNR	DNR
5	15	60	0	16	18	0.0265	0.4086	0.4147	DNR	DNR	DNR
6	5	40	0	19	24	0.9193	0.8686	0.8694	DNR	DNR	DNR
6	5	80	0	55	58	0.5082	0.1965	0.1992	DNR	DNR	DNR
6	10	20	0	111	117	0.0161	0.0091	0.0097	DNR	DNR	DNR
6	10	40	0	62	72	0.2833	0.3891	0.3907	DNR	DNR	DNR
6	10	60	0	42	47	0.1369	0.0392	0.0422	DNR	DNR	DNR
6	10	80	0	51	57	0.3393	0.4998	0.5013	DNR	DNR	DNR
6	15	20	0	22	23	0.8871	0.8464	0.8473	DNR	DNR	DNR
6	15	40	0	18	23	0.9592	0.8126	0.8138	DNR	DNR	DNR
6	15	60	0	18	21	0.1159	0.2557	0.2630	DNR	DNR	DNR
6	15	80	0	25	25	0.9896	0.8861	0.8867	DNR	DNR	DNR
7	5	60	0	47	73	0.6485	0.1894	0.1919	DNR	DNR	DNR
7	5	80	0	143	231	0.9707	0.3753	0.3759	DNR	DNR	DNR
7	10	20	0	55	136	0.0644	0.5915	0.5921	DNR	DNR	DNR
7	10	40	0	129	189	0.5809	0.4635	0.4640	DNR	DNR	DNR
7	10	60	0	167	199	0.6478	0.7768	0.7770	DNR	DNR	DNR
7	10	80	0	259	346	0.5969	0.4232	0.4235	DNR	DNR	DNR
7	10	100	0	13	20	0.8874	0.5906	0.5944	DNR	DNR	DNR
7	15	20	0	27	36	0.5563	0.9017	0.9021	DNR	DNR	DNR
7	15	40	0	47	76	0.0035	0.2769	0.2790	DNR	DNR	DNR
7	15	60	0	75	88	0.9792	0.8886	0.8888	DNR	DNR	DNR
7	15	80	0	74	97	0.9964	0.9079	0.9080	DNR	DNR	DNR
7	20	60	0	14	24	0.8439	0.4591	0.4639	DNR	DNR	DNR
7	20	80	0	20	27	0.0434	0.0754	0.0822	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.12: Shenyang-Fushun

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
2	5	20	0	20	11	0.6791	0.7903	0.7922	DNR	DNR	DNR
3	5	20	0	52	18	0.2609	0.5721	0.5739	DNR	DNR	DNR
3	5	40	0	32	17	0.6064	0.2441	0.2499	DNR	DNR	DNR
3	5	80	0	17	14	0.6253	0.7506	0.7529	DNR	DNR	DNR
4	5	20	0	48	31	0.8645	0.5686	0.5703	DNR	DNR	DNR
4	5	40	0	54	30	0.0015	0.0054	0.0067	DNR	DNR	DNR
4	5	60	0	29	23	0.2930	0.4554	0.4589	DNR	DNR	DNR
4	5	80	0	57	43	0.1626	0.3612	0.3634	DNR	DNR	DNR
4	10	60	0	18	12	0.6927	0.3610	0.3688	DNR	DNR	DNR
4	10	80	0	25	16	0.1991	0.0396	0.0464	DNR	DNR	DNR
5	5	20	0	46	38	0.2011	0.2416	0.2450	DNR	DNR	DNR
5	5	40	0	44	27	0.0603	0.2445	0.2485	DNR	DNR	DNR
5	5	60	0	38	23	0.4710	0.0860	0.0912	DNR	DNR	DNR
5	5	80	0	129	89	0.0000	0.0000	0.0000	R	R	R
5	10	20	0	29	14	0.8690	0.3804	0.3855	DNR	DNR	DNR
5	10	40	0	40	22	0.3137	0.8955	0.8959	DNR	DNR	DNR
5	10	60	0	36	23	0.5988	0.9515	0.9517	DNR	DNR	DNR
5	10	80	0	51	28	0.0005	0.0091	0.0109	R	DNR	DNR
6	5	20	0	70	70	0.4429	0.6010	0.6018	DNR	DNR	DNR
6	5	40	0	61	45	0.0637	0.0158	0.0176	DNR	DNR	DNR
6	5	60	0	53	40	0.0192	0.0170	0.0191	DNR	DNR	DNR
6	5	80	0	160	124	0.0000	0.0000	0.0000	R	R	R
6	10	20	0	50	36	0.8603	0.9891	0.9892	DNR	DNR	DNR
6	10	40	0	62	54	0.0383	0.1368	0.1396	DNR	DNR	DNR
6	10	60	0	81	61	0.3513	0.8743	0.8745	DNR	DNR	DNR
6	10	80	0	169	114	0.0001	0.0000	0.0000	R	R	R
6	15	40	0	22	19	0.7941	0.2323	0.2396	DNR	DNR	DNR
6	15	60	0	26	18	0.1820	0.0982	0.1057	DNR	DNR	DNR
7	5	20	0	102	123	0.0113	0.1624	0.1638	DNR	DNR	DNR
7	5	40	0	96	97	0.2004	0.0385	0.0398	DNR	DNR	DNR
7	5	60	0	123	109	0.0045	0.0179	0.0188	DNR	DNR	DNR
7	5	80	0	264	207	0.0000	0.0000	0.0000	R	R	R
7	10	20	0	66	79	0.7460	0.5636	0.5645	DNR	DNR	DNR
7	10	40	0	161	163	0.0007	0.0079	0.0083	DNR	DNR	DNR
7	10	60	0	169	149	0.0013	0.0312	0.0319	DNR	DNR	DNR
7	10	80	0	283	213	0.0000	0.0000	0.0000	R	R	R
7	15	40	0	23	18	0.0725	0.3446	0.3504	DNR	DNR	DNR
7	15	60	0	26	16	0.3860	0.4585	0.4628	DNR	DNR	DNR
7	15	80	0	30	23	0.0637	0.0529	0.0584	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.13: Yinchuan-Shizuishan

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
5	5	40	0.0	25	11	0.6907	0.4161	0.4217	DNR	DNR	DNR
6	5	40	0	39	14	0.9683	0.8619	0.8626	DNR	DNR	DNR
7	5	20	0	63	31	0.9496	0.5733	0.5747	DNR	DNR	DNR
7	5	40	0	334	115	0.4478	0.6296	0.6299	DNR	DNR	DNR
7	5	60	0	348	205	0.0434	0.1240	0.1246	DNR	DNR	DNR
7	5	80	0	593	416	0.0425	0.0007	0.0007	DNR	R	R
7	5	100	0	69	41	0.8509	0.5302	0.5315	DNR	DNR	DNR
7	10	20	0	22	14	0.419863	0.642193	0.645155	DNR	DNR	DNR
7	10	40	0	154	68	0.099932	0.397501	0.39842	DNR	DNR	DNR
7	10	60	0	289	145	0.000108	0.001937	0.002064	R	R	DNR
7	10	80	0	521	270	0.002098	0.009442	0.009618	DNR	DNR	DNR
7	10	100	0	40	16	0.211188	0.099758	0.105567	DNR	DNR	DNR
7	15	60	0	35	14	0.460835	0.844428	0.845274	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.14: Xian-Xianyang I

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
1	5	40	0	153	152	0.0357	0.4504	0.4510	DNR	DNR	DNR
1	5	60	0	91	82	0.0932	0.2174	0.2191	DNR	DNR	DNR
1	5	80	0	66	62	0.0002	0.0000	0.0000	R	R	R
1	10	40	0	22	23	0.9321	0.9344	0.9348	DNR	DNR	DNR
1	10	60	0	16	15	0.8112	0.9337	0.9342	DNR	DNR	DNR
1	10	80	0	32	29	0.0032	0.0001	0.0003	DNR	R	R
2	5	40	0	96	88	0.0744	0.0230	0.0242	DNR	DNR	DNR
2	5	60	0	119	104	0.3121	0.6448	0.6452	DNR	DNR	DNR
2	5	80	0	54	40	0.0120	0.0011	0.0016	DNR	DNR	DNR
2	10	40	0	18	17	0.5377	0.4727	0.4778	DNR	DNR	DNR
2	10	60	0	37	33	0.0425	0.0171	0.0199	DNR	DNR	DNR
2	10	80	0	66	59	0.0005	0.0015	0.0019	DNR	DNR	DNR
2	10	100	0	23	21	0.0000	0.0000	0.0000	R	R	R
2	15	80	0	18	16	0.0018	0.0000	0.0002	DNR	R	R
3	5	40	0	120	93	0.0000	0.0000	0.0000	R	R	R
3	5	60	0	121	94	0.2578	0.0772	0.0786	DNR	DNR	DNR
3	5	80	0	99	71	0.1335	0.2693	0.2709	DNR	DNR	DNR
3	10	40	0	27	19	0.4742	0.2610	0.2671	DNR	DNR	DNR
3	10	60	0	39	35	0.4048	0.1966	0.2007	DNR	DNR	DNR
3	10	80	0	65	49	0.0047	0.0063	0.0073	DNR	DNR	DNR
3	10	100	0	13	13	0.0280	0.0943	0.1073	DNR	DNR	DNR
3	15	80	0	19	13	0.1115	0.7014	0.7041	DNR	DNR	DNR
3	15	100	0	16	15	0.0001	0.0000	0.0001	R	R	R
4	5	40	0	97	76	0.1567	0.0143	0.0153	DNR	DNR	DNR
4	5	60	0	127	93	0.2021	0.0565	0.0578	DNR	DNR	DNR
4	5	80	0	104	77	0.0778	0.8732	0.8734	DNR	DNR	DNR
4	5	100	0	17	14	0.1909	0.5901	0.5942	DNR	DNR	DNR
4	10	40	0	30	26	0.6331	0.2689	0.2738	DNR	DNR	DNR
4	10	60	0	67	50	0.1752	0.5326	0.5338	DNR	DNR	DNR
4	10	80	0	86	67	0.0012	0.1289	0.1310	DNR	DNR	DNR
4	10	100	0	14	12	0.0300	0.2191	0.2310	DNR	DNR	DNR
5	5	40	0	56	45	0.0429	0.0013	0.0017	DNR	DNR	DNR
5	5	60	0	92	67	0.0001	0.3793	0.3807	R	DNR	DNR
5	5	80	0	116	88	0.0058	0.2991	0.3004	DNR	DNR	DNR
5	5	100	0	13	11	0.1208	0.4876	0.4949	DNR	DNR	DNR
5	10	40	0	21	19	0.9074	0.6547	0.6573	DNR	DNR	DNR
5	10	60	0	42	39	0.2323	0.3831	0.3857	DNR	DNR	DNR
5	10	80	0	74	60	0.0018	0.5990	0.5999	DNR	DNR	DNR
5	10	100	0	20	16	0.0461	0.7430	0.7450	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.15: Xian-Xianyang II

VSB	SPD	Temp	PCP	Subs 1	Subs 2	KS	Z	T	KS	Z	T
6	5	40	0	41	35	0.0491	0.0184	0.0211	DNR	DNR	DNR
6	5	60	0	76	65	0.1816	0.4096	0.4110	DNR	DNR	DNR
6	5	80	0	128	105	0.0252	0.8945	0.8947	DNR	DNR	DNR
6	5	100	0	46	37	0.1707	0.5414	0.5431	DNR	DNR	DNR
6	10	40	0	39	40	0.2456	0.4660	0.4682	DNR	DNR	DNR
6	10	60	0	71	60	0.1010	0.8141	0.8145	DNR	DNR	DNR
6	10	80	0	90	75	0.0009	0.0174	0.0186	DNR	DNR	DNR
6	10	100	0	47	41	0.0013	0.1373	0.1410	DNR	DNR	DNR
6	15	80	0	23	18	0.0963	0.3057	0.3120	DNR	DNR	DNR
7	5	60	0	28	18	0.9672	0.5727	0.5756	DNR	DNR	DNR
7	5	80	0	99	59	0.0298	0.0374	0.0390	DNR	DNR	DNR
7	5	100	0	70	50	0.5321	0.7907	0.7912	DNR	DNR	DNR
7	10	60	0	26	15	0.4083	0.2688	0.2756	DNR	DNR	DNR
7	10	80	0	73	50	0.0375	0.1524	0.1550	DNR	DNR	DNR
7	10	100	0	71	45	0.3448	0.2232	0.2257	DNR	DNR	DNR

Note: KS denotes the two-sample Kolmogorov-Smirnov test, Z denotes the test based on the normal approximation, T denotes the test based on the t-distribution. P-values are reported for the three tests. DNR and R denote 'do not reject' and 'reject', respectively. We correct for multiple testing using $\alpha = 0.05$, see Appendix B for details.

Table SA.16: Robustness Checks for PM_{10} McCrary Test: Discontinuity at $c = 0.1$ I

	a			$\hat{P}(PM_{10} > c)$	$\#(PM_{10} > c)$		a			$\hat{P}(PM_{10} > c)$	$\#(PM_{10} > c)$
	10	15	20				10	15	20		
Shenzhen	NaN	-2.11	-2.31	0.002	8	Zhenjiang	-0.18	-0.31	-0.32	0.058	160
Deyang	-1.79	-1.62	-1.26	0.008	21	Changzhi	0.27	-0.26	-0.14	0.153	480
Zhengzhou	-1.44	-1.50	-1.05	0.057	207	Baoji	-0.13	-0.24	0.72	0.091	286
Anshan	-0.87	-1.43	-1.99	0.074	206	Mudanjiang	0.46	-0.22	-0.43	0.047	148
Qiqihaer	-1.29	-1.34	-1.60	0.022	69	Shaoxing	-0.29	-0.20	-0.09	0.050	157
Jiujiang	-0.74	-1.28	-1.57	0.023	65	Guiyang	-1.10	-0.20	0.79	0.013	46
Mianyang	-1.20	-1.22	-0.65	0.016	51	Maanshan	0.22	-0.19	-0.12	0.016	44
Xianyang	-1.43	-1.21	1.57	0.123	385	Shantou	NaN	-0.17	-0.22	0.001	3
Wuhu	-1.08	-1.08	-1.21	0.006	17	Kunming	NaN	-0.16	1.22	0.001	3
Nanning	-0.86	-1.00	-0.85	0.003	12	Dalian	-0.32	-0.16	-0.08	0.016	58
Jinan	-0.65	-0.99	-1.34	0.113	408	Rizhao	-0.14	-0.13	-0.48	0.005	17
Benxi	-0.96	-0.77	-0.79	0.050	137	Baotou	-0.64	-0.13	0.98	0.207	650
Luzhou	-0.77	-0.74	-0.83	0.100	316	Jinchang	-0.38	-0.13	-0.12	0.058	183
Datong	-0.32	-0.69	-0.22	0.168	528	Lianyungang	0.01	-0.13	-0.09	0.053	191
Xining	-0.47	-0.66	0.30	0.099	359	Changde	0.15	-0.07	0.16	0.037	116
Zibo	-1.16	-0.65	0.33	0.024	76	Lasa	-0.25	0.00	-0.17	0.011	40
Yinchuan	0.14	-0.58	-0.77	0.049	177	Weifang	0.31	0.00	-0.10	0.014	39
Changchun	-0.36	-0.58	-0.82	0.032	116	Liuzhou	0.00	0.06	0.30	0.004	11
Suzhou	-0.08	-0.53	-0.94	0.071	257	Yueyang	0.81	0.07	-0.11	0.104	327
Zunyi	-0.49	-0.51	0.15	0.032	102	Baoding	0.29	0.17	-0.50	0.058	181
Changzhou	0.11	-0.49	-0.95	0.046	145	Lanzhou	-0.09	0.18	0.54	0.265	958
Hefei	1.05	-0.45	-1.11	0.072	261	Yangquan	-0.33	0.20	0.60	0.089	245
Zigong	-0.38	-0.36	-0.50	0.005	10	Xiangtan	0.81	0.22	-0.12	0.127	398
Zaozhuang	0.20	-0.35	-0.44	0.044	137	Shenyang	1.05	0.25	0.08	0.110	397
Fuzhou	0.45	-0.34	-0.61	0.006	21	Wuxi	0.47	0.26	0.50	0.039	124

This table reports the McCrary t-statistic for each city using the PM_{10} concentration data with cutoff for discontinuity at $c = 0.1$.

Notes: $a = h/\hat{b}$, h is the bandwidth and $\hat{b} = 2\hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is the standard deviation of the pollutant concentration. For more details, please see the appendix in the paper.

Table SA.17: Robustness Checks for PM_{10} McCrary Test: Discontinuity at $c = 0.1$ II

	a						a				
	10	15	20	$\hat{P}(PM_{10} > c)$	$\#(PM_{10} > c)$		10	15	20	$\hat{P}(PM_10 > c)$	$\#(PM_10 > c)$
'Taian'	0.04	0.30	0.07	0.010	31	'Panzhihua'	1.91	1.53	1.32	0.078	245
'Hangzhou'	1.10	0.30	0.15	0.082	298	'Xuzhou'	1.49	1.60	2.00	0.164	515
'Wenzhou'	NaN	0.38	0.34	0.004	14	'Pingdingshan'	1.56	1.60	1.77	0.137	429
'Shanghai'	0.17	0.38	0.47	0.052	189	'Yangzhou'	1.94	1.70	1.31	0.068	213
'Xian'	0.74	0.46	0.26	0.123	445	'Chifeng'	1.47	1.79	2.41	0.082	258
'Yichang'	0.43	0.52	0.89	0.034	95	'Ningbo'	1.56	1.79	1.80	0.033	116
'Changsha'	1.23	0.53	-0.14	0.106	384	'Qingdao'	1.91	1.80	1.53	0.033	119
'Jining'	0.45	0.54	0.64	0.014	39	'Anyang'	2.06	1.92	1.93	0.136	428
'Guangzhou'	0.79	0.55	0.67	0.018	65	'Zhuzhou'	2.39	1.95	2.45	0.088	277
'Nantong'	0.08	0.57	0.57	0.056	201	'Fushun'	1.47	1.97	2.39	0.067	184
'Sanmenxia'	0.84	0.63	0.88	0.072	199	'Tongchuan'	2.33	2.04	1.72	0.121	380
'Luoyang'	1.51	0.69	0.05	0.160	504	'Handan'	2.37	2.12	1.79	0.079	218
'Yibin'	0.74	0.82	0.48	0.045	140	'Taiyuan'	2.17	2.21	2.38	0.169	611
'Shijiazhuang'	1.04	0.85	0.43	0.151	548	'Shaoguan'	2.10	2.24	2.27	0.017	54
'Haerbin'	0.70	0.91	1.17	0.074	268	'Tangshan'	2.46	2.63	2.59	0.045	142
'Shizuishan'	1.24	0.96	1.28	0.041	103	'Wuhan'	2.81	2.68	2.37	0.107	389
'Chongqing'	1.07	0.97	1.15	0.119	422	'Quanzhou'	2.62	2.69	2.56	0.015	48
'Yantai'	0.42	0.99	1.08	0.008	30	'Wulumuqi'	2.86	3.98	4.69	0.200	724
'Nanchang'	1.22	1.03	0.94	0.043	156	'Qinhuangdao'	2.56	5.17	-0.51	0.005	18
'Huzhou'	1.03	1.08	1.15	0.020	64	'Linfen'	3.37	5.49	4.43	0.238	748
'Tianjin'	1.29	1.12	0.54	0.092	334	'Beihai'	NaN	NaN	NaN	0.000	1
'Jinzhou'	0.66	1.12	1.46	0.049	136	'Guilin'	NaN	NaN	NaN	0.000	1
'Beijing'	1.61	1.17	1.72	0.197	712	'Haikou'	NaN	NaN	NaN	0.000	0
'Weinan'	1.28	1.19	1.63	0.152	420	'Kelamayi'	NaN	NaN	NaN	0.004	13
'Nanjing'	1.36	1.21	0.94	0.094	340	'Nanchong'	NaN	NaN	NaN	0.001	1
'Jiaozuo'	0.86	1.21	1.90	0.073	203	'Qujing'	NaN	NaN	NaN	0.000	0
'Huhehaote'	1.64	1.29	1.80	0.076	275	'Xiamen'	NaN	NaN	NaN	0.002	6
'Chengdou'	1.23	1.35	1.51	0.065	233	'Yuxi'	NaN	NaN	NaN	0.000	0
'Kaifeng'	1.85	1.35	1.49	0.142	448	'Zhanjiang'	NaN	NaN	NaN	0.000	0
'Yanan'	1.51	1.43	1.03	0.083	262	'Zhuhai'	NaN	NaN	NaN	0.000	0
'Jilin'	1.65	1.45	1.37	0.044	123						

This table reports the McCrary t-statistic for each city using the PM_{10} concentration data with cutoff for discontinuity at $c = 0.1$.

Notes: $a = h/\hat{b}$, h is the bandwidth and $\hat{b} = 2\hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is the standard deviation of the pollutant concentration. For more details, please see the appendix in the paper. $\hat{P}(PM_{10} > c)$ and $\#(PM_{10} > c)$ denote the empirical probability and the number of observations, respectively, that PM_{10} is greater than the cutoff c specified above.

Table SA.18: Robustness Checks for PM_{10} McCrary Test: Discontinuity at $c = 0.2$ I

	10	15	20	P($PM_{10} > c$)	$\sum(PM_10 > c)$		10	15	20	$P(PM_{10} > c)$	$\sum(PM_{10} > c)$
'Shenzhen'	NaN	-2.11	-2.31	0.002	8	'Zhenjiang'	-0.18	-0.31	-0.32	0.058	160
'Deyang'	-1.79	-1.62	-1.26	0.008	21	'Changzhi'	0.27	-0.26	-0.14	0.153	480
'Zhengzhou'	-1.44	-1.50	-1.05	0.057	207	'Baoji'	-0.13	-0.24	0.72	0.091	286
'Anshan'	-0.87	-1.43	-1.99	0.074	206	'Mudanjiang'	0.46	-0.22	-0.43	0.047	148
'Qiqihaer'	-1.29	-1.34	-1.60	0.022	69	'Shaoxing'	-0.29	-0.20	-0.09	0.050	157
'Jiujiang'	-0.74	-1.28	-1.57	0.023	65	'Guiyang'	-1.10	-0.20	0.79	0.013	46
'Mianyang'	-1.20	-1.22	-0.65	0.016	51	'Maanshan'	0.22	-0.19	-0.12	0.016	44
'Xianyang'	-1.43	-1.21	1.57	0.123	385	'Shantou'	NaN	-0.17	-0.22	0.001	3
'Wuhu'	-1.08	-1.08	-1.21	0.006	17	'Kunming'	NaN	-0.16	1.22	0.001	3
'Nanning'	-0.86	-1.00	-0.85	0.003	12	'Dalian'	-0.32	-0.16	-0.08	0.016	58
'Jinan'	-0.65	-0.99	-1.34	0.113	408	'Rizhao'	-0.14	-0.13	-0.48	0.005	17
'Benxi'	-0.96	-0.77	-0.79	0.050	137	'Baotou'	-0.64	-0.13	0.98	0.207	650
'Luzhou'	-0.77	-0.74	-0.83	0.100	316	'Jinchang'	-0.38	-0.13	-0.12	0.058	183
'Datong'	-0.32	-0.69	-0.22	0.168	528	'Lianyungang'	0.01	-0.13	-0.09	0.053	191
'Xining'	-0.47	-0.66	0.30	0.099	359	'Changde'	0.15	-0.07	0.16	0.037	116
'Zibo'	-1.16	-0.65	0.33	0.024	76	'Lasa'	-0.25	0.00	-0.17	0.011	40
'Yinchuan'	0.14	-0.58	-0.77	0.049	177	'Weifang'	0.31	0.00	-0.10	0.014	39
'Changchun'	-0.36	-0.58	-0.82	0.032	116	'Liuzhou'	0.00	0.06	0.30	0.004	11
'Suzhou'	-0.08	-0.53	-0.94	0.071	257	'Yueyang'	0.81	0.07	-0.11	0.104	327
'Zunyi'	-0.49	-0.51	0.15	0.032	102	'Baoding'	0.29	0.17	-0.50	0.058	181
'Changzhou'	0.11	-0.49	-0.95	0.046	145	'Lanzhou'	-0.09	0.18	0.54	0.265	958
'Hefei'	1.05	-0.45	-1.11	0.072	261	'Yangquan'	-0.33	0.20	0.60	0.089	245
'Zigong'	-0.38	-0.36	-0.50	0.005	10	'Xiangtan'	0.81	0.22	-0.12	0.127	398
'Zaozhuang'	0.20	-0.35	-0.44	0.044	137	'Shenyang'	1.05	0.25	0.08	0.110	397
'Fuzhou'	0.45	-0.34	-0.61	0.006	21	'Wuxi'	0.47	0.26	0.50	0.039	124

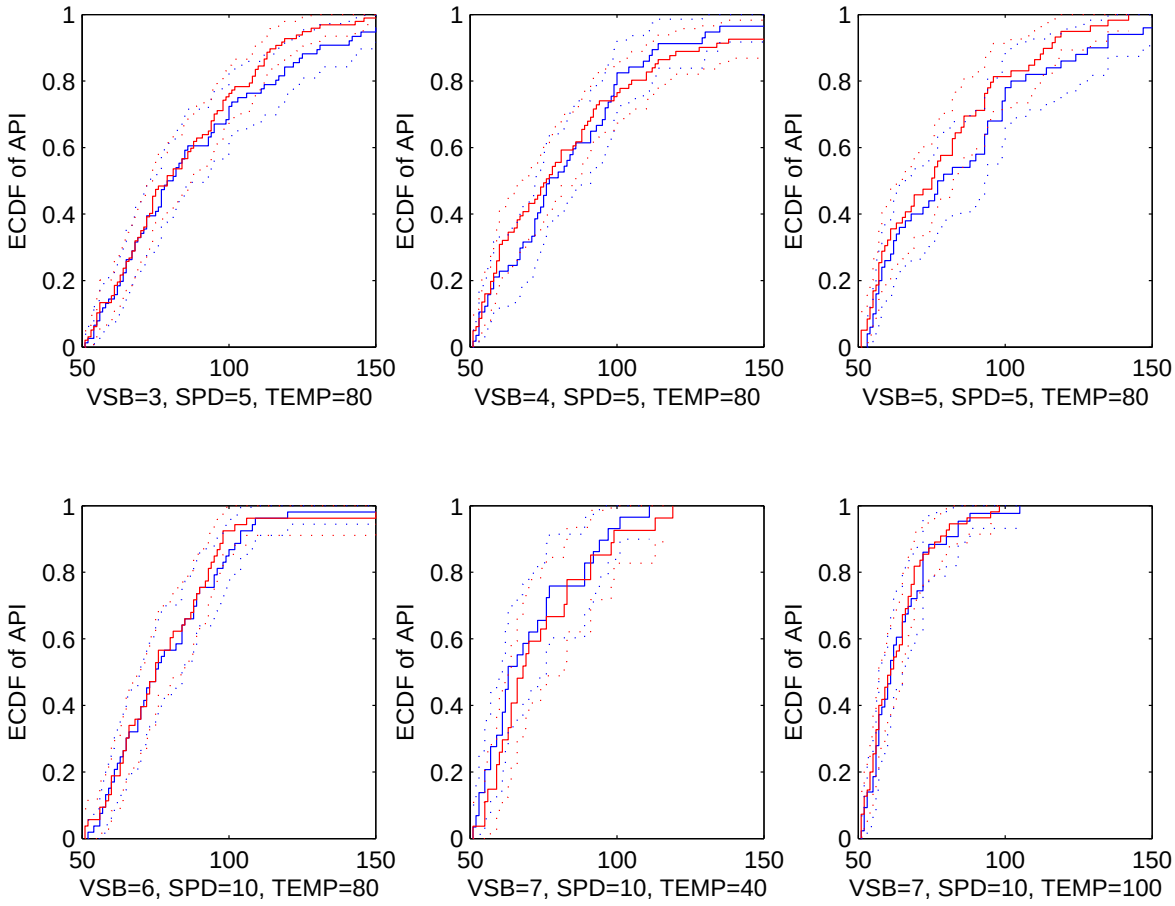


Figure SA.1: **Panel Matching Approach: Zhenjiang-Yangzhou.** The red lines represent Zhenjiang, the blue Yangzhou. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

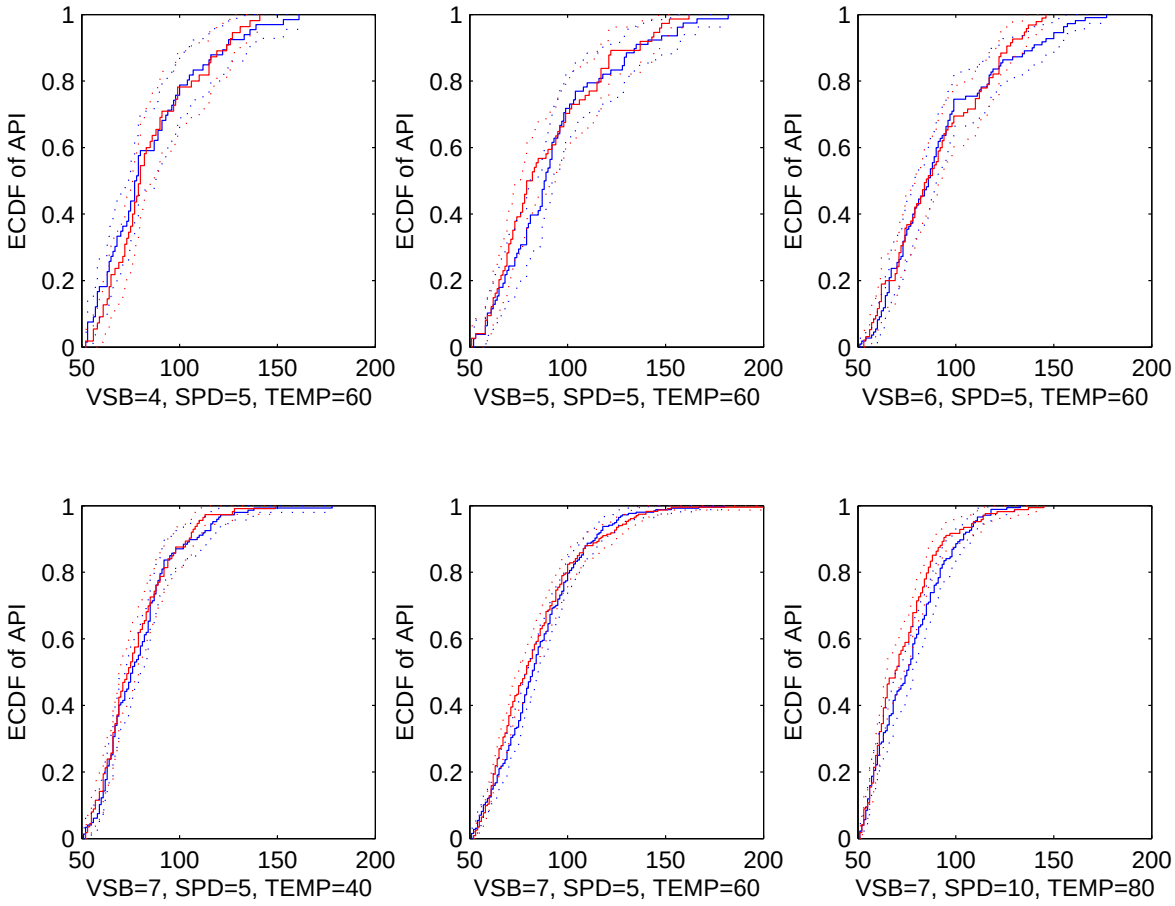


Figure SA.2: **Panel Matching Approach: Changzhou-Wuxi.** The red lines represent Changzhou, the blue Wuxi. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

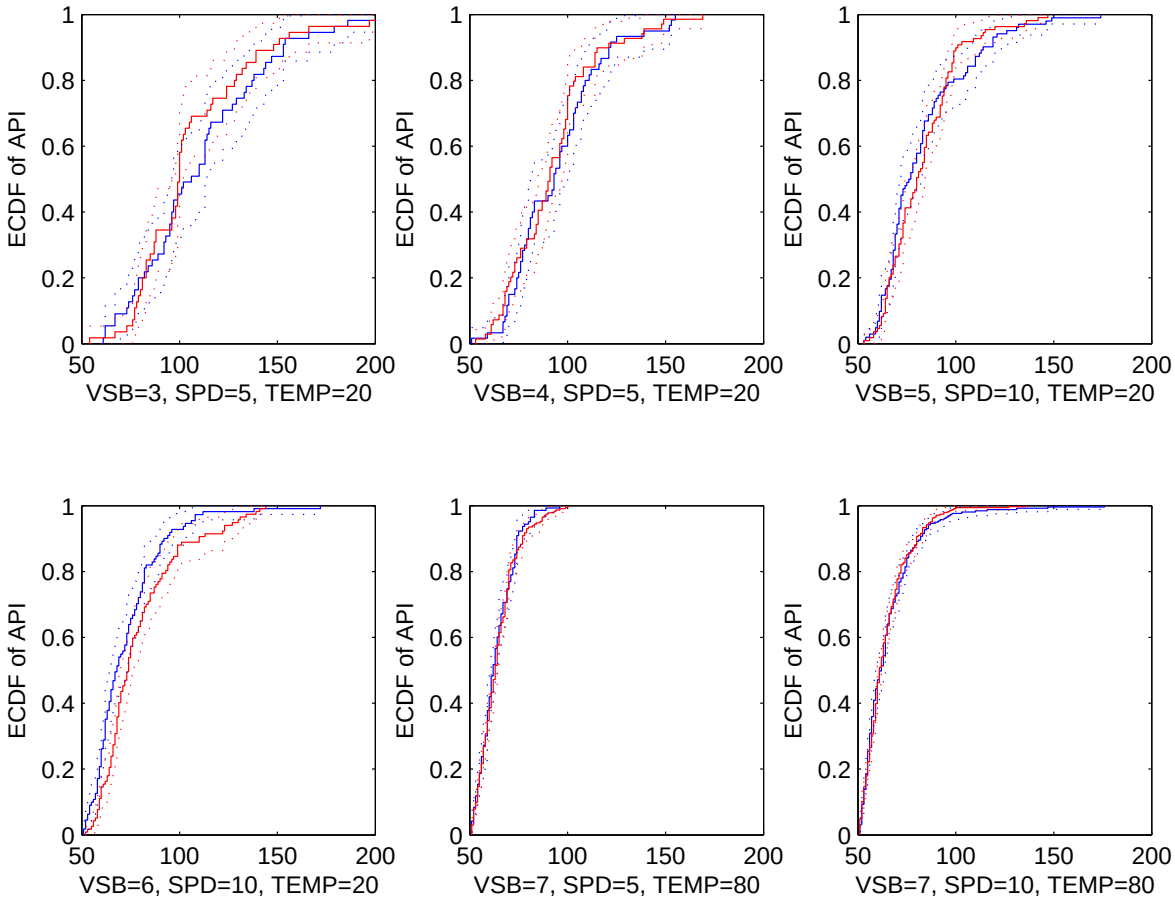


Figure SA.3: **Panel Matching Approach: Jilin-Changchun.** The red lines represent Jilin, the blue Changchun. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

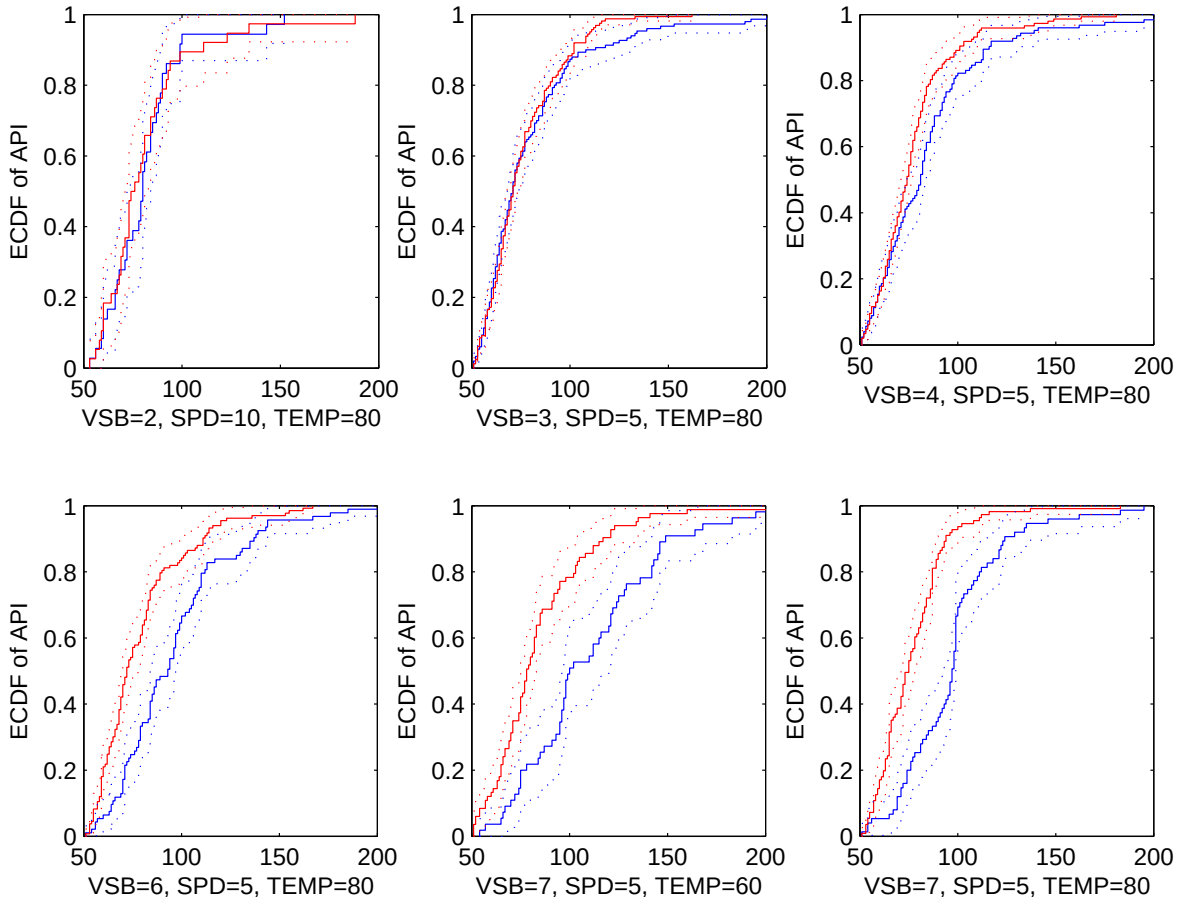


Figure SA.4: **Panel Matching Approach: Kaifeng-Zhengzhou.** The red lines represent Kaifeng, the blue Zhengzhou. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

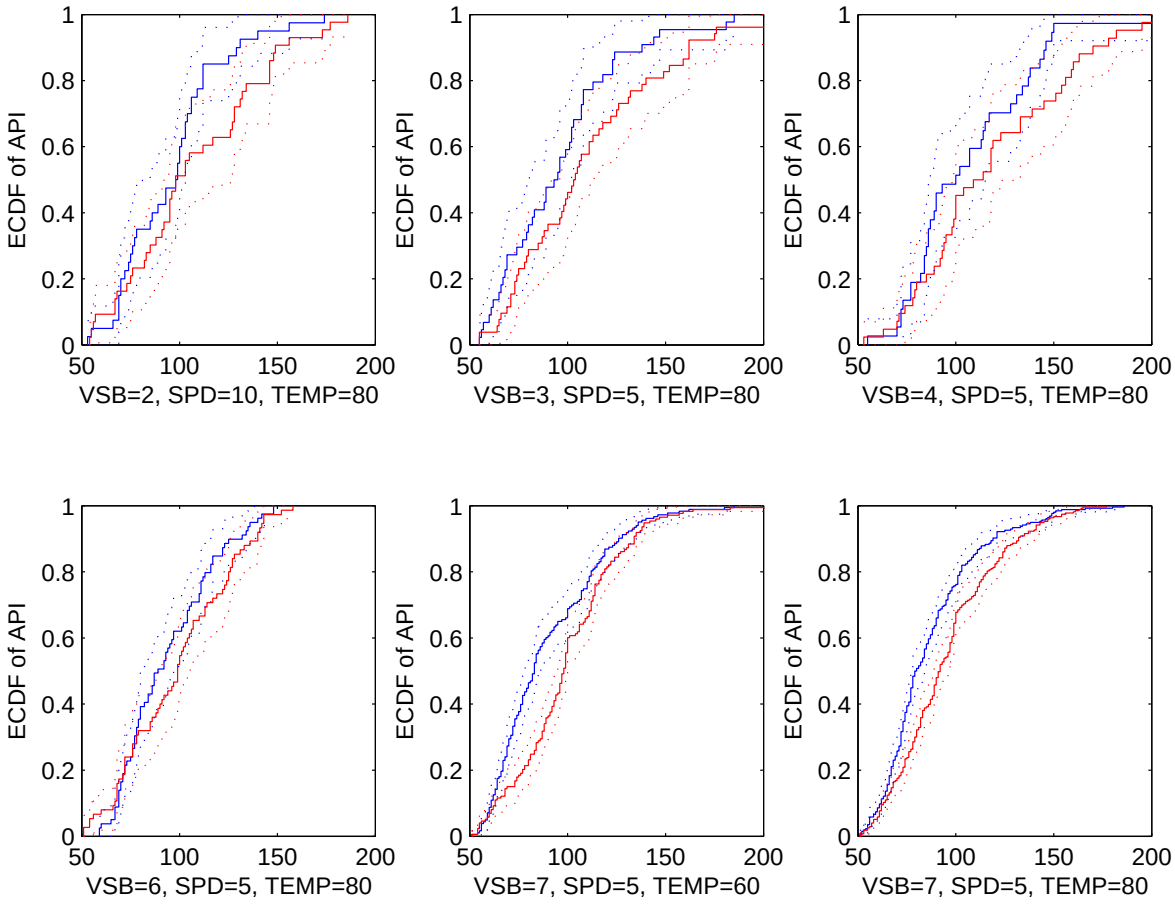


Figure SA.5: **Panel Matching Approach: Zhuzhou-Xiangtan.** The red lines represent Zhuzhou, the blue Xiangtan. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

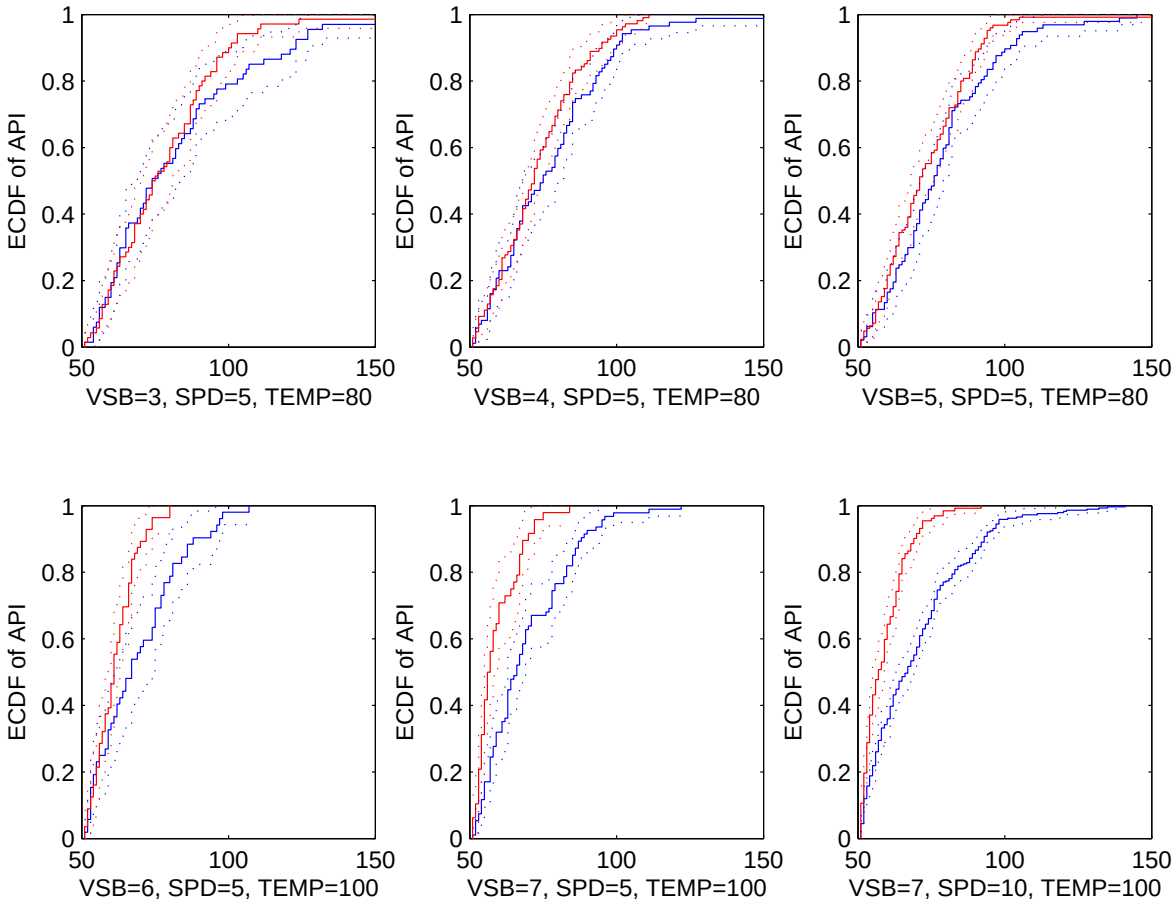


Figure SA.6: **Panel Matching Approach: Quanzhou-Xiamen.** The red lines represent Quanzhou, the blue Xiamen. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

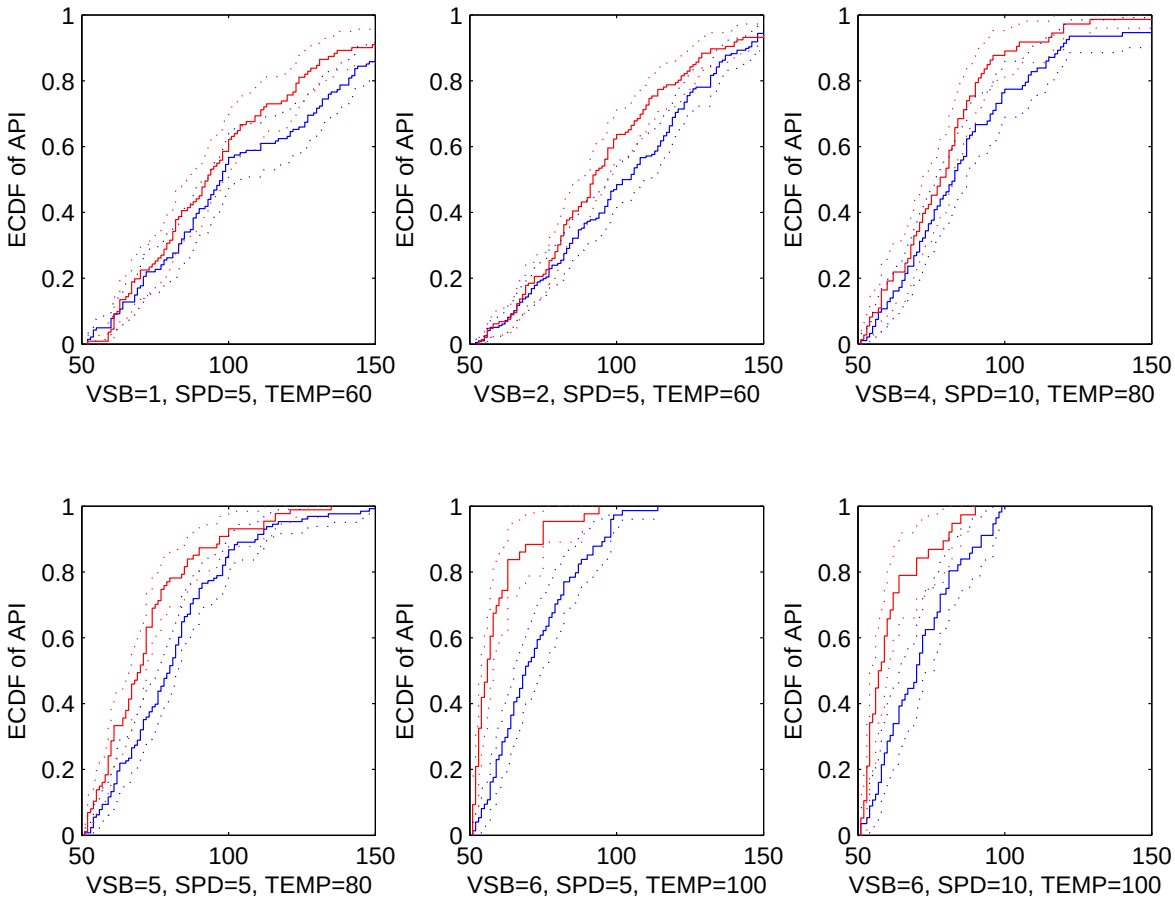


Figure SA.7: **Panel Matching Approach: Hangzhou-Shaoxing.** The red lines represent Hangzhou, the blue Shaoxing. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

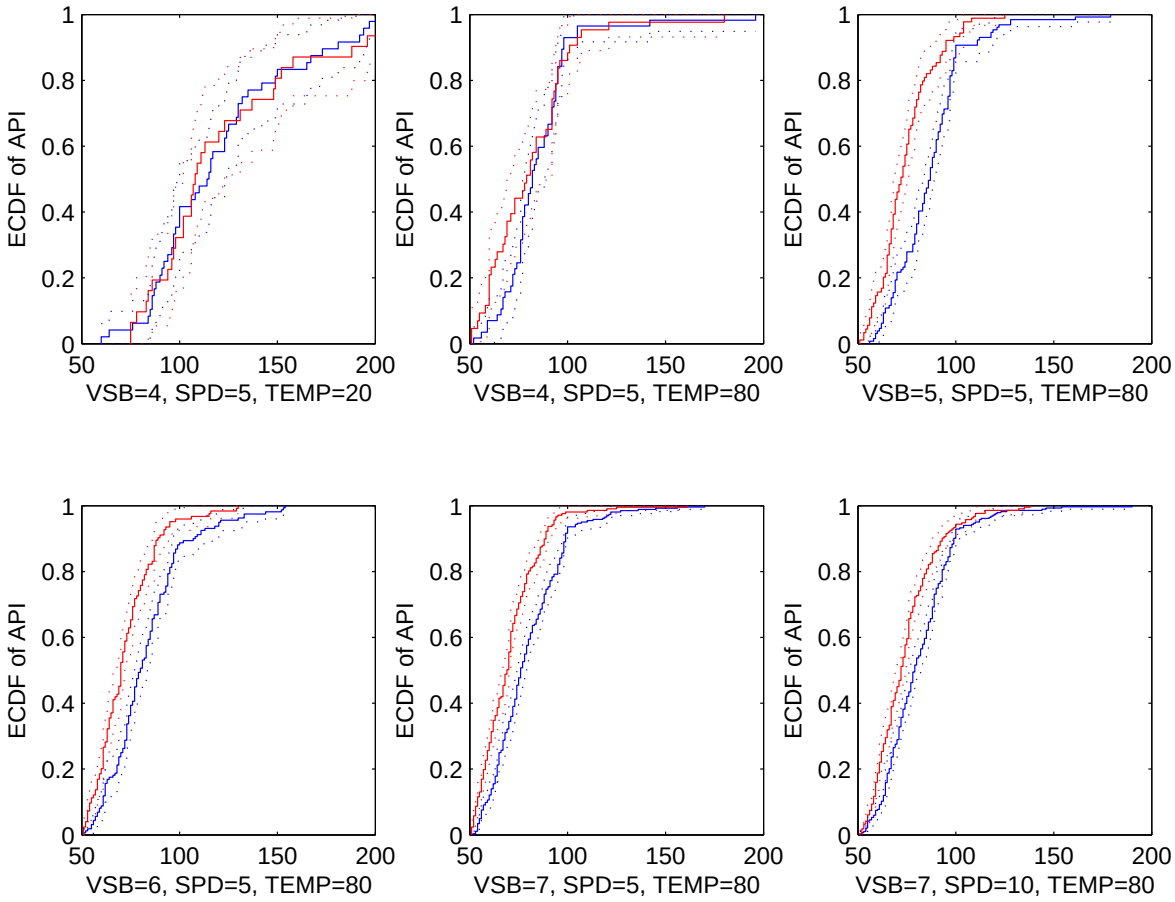


Figure SA.8: **Panel Matching Approach: Shenyang-Fushun.** The red lines represent Shenyang, the blue Fushun. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

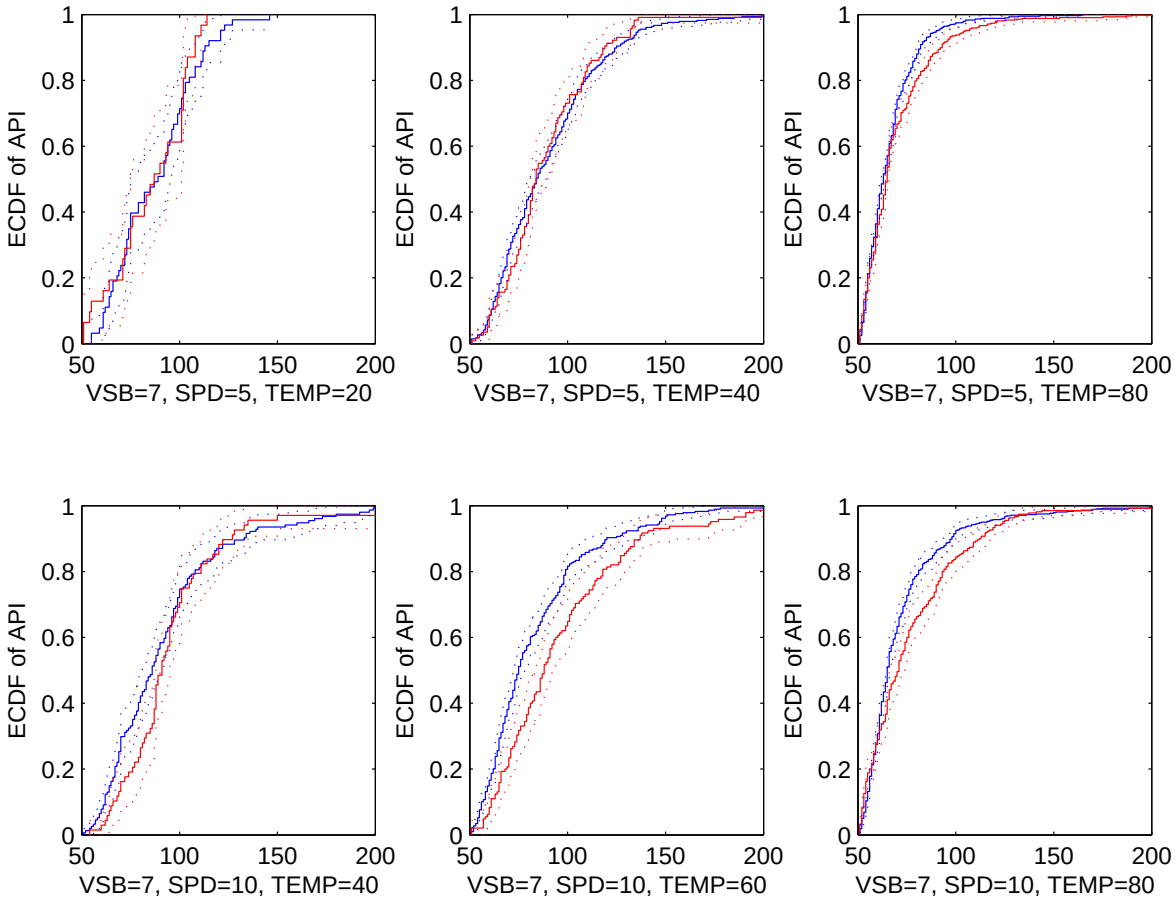


Figure SA.9: **Panel Matching Approach: Yinchuan-Shizuishan.** The red lines represent Yinchuan, the blue Shizuishan. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

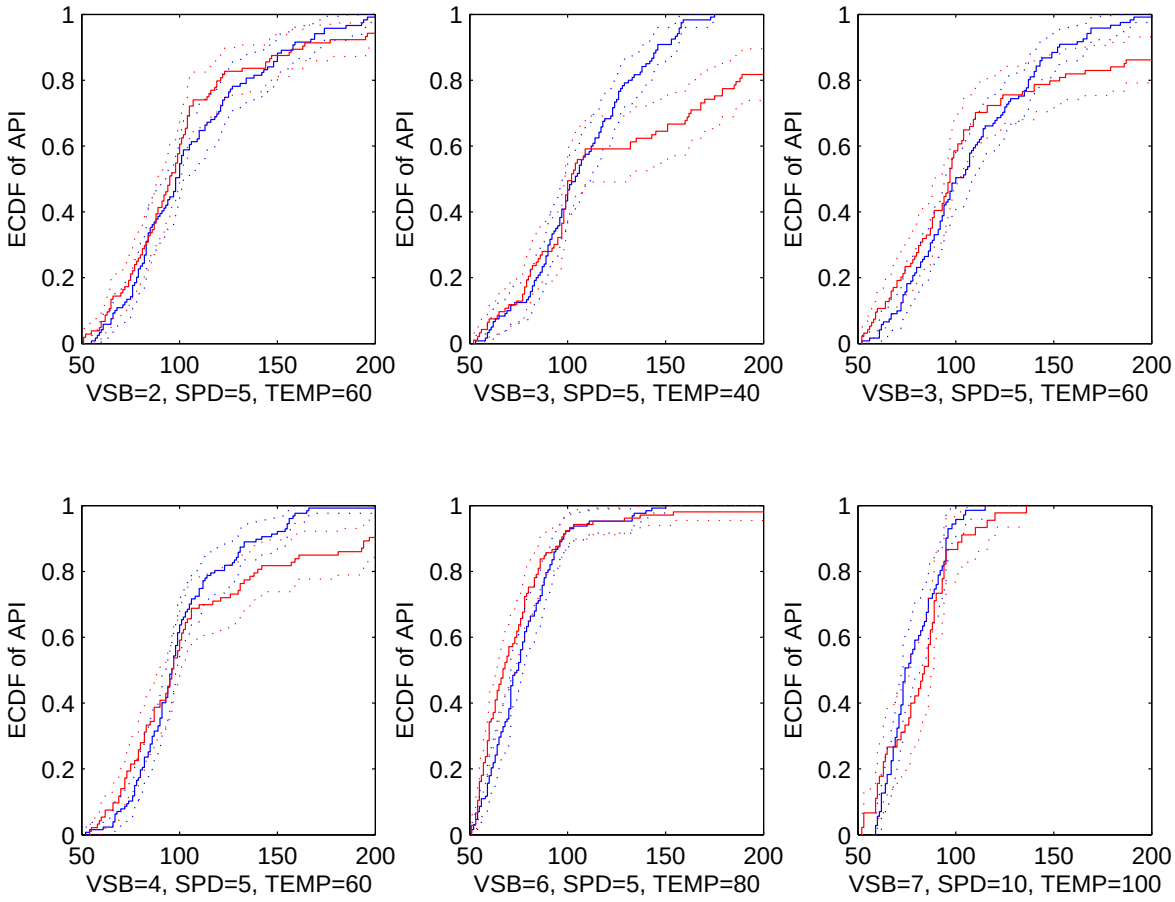


Figure SA.10: **Panel Matching Approach: Xian-Xianyang.** The red lines represent Xian, the blue Xianyang. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

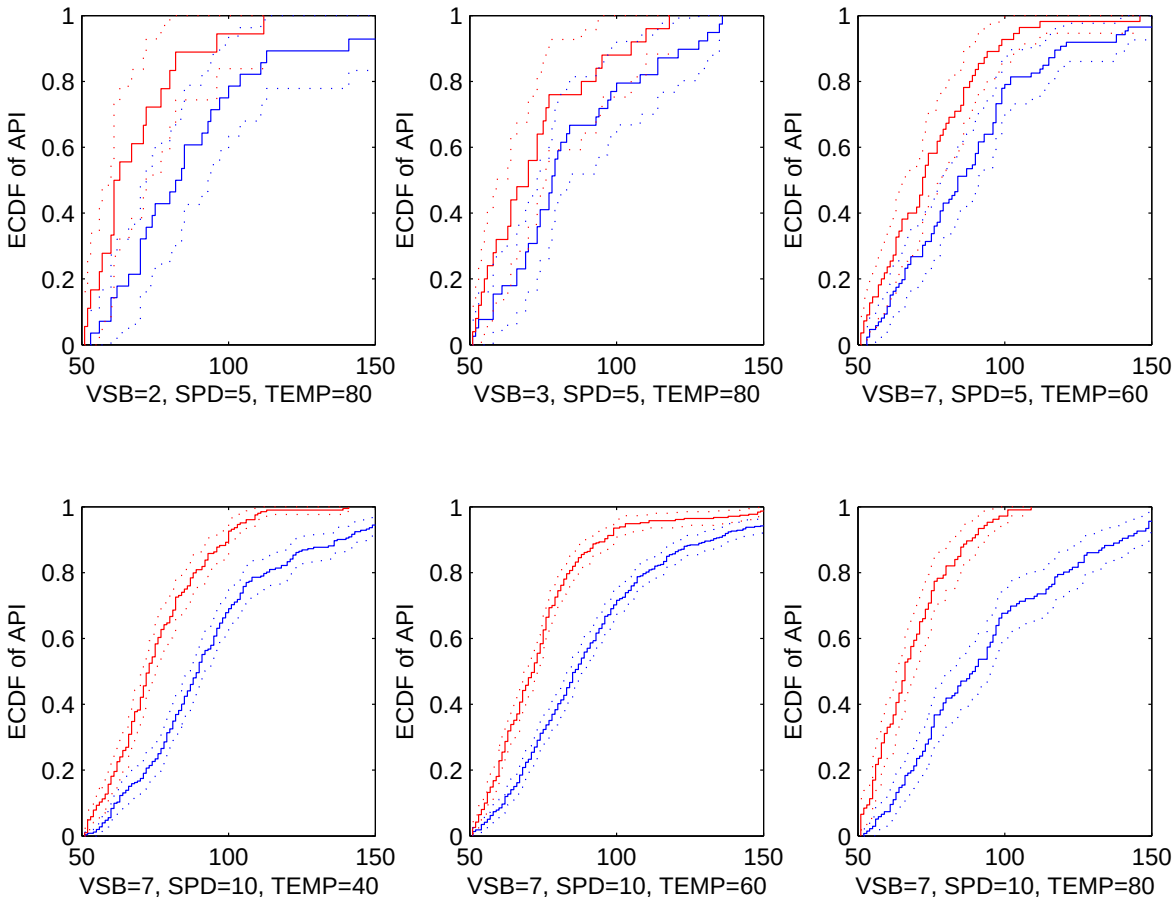


Figure SA.11: **Panel Matching Approach: Jinan-Taian.** The red lines represent Jinan, the blue Taian. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

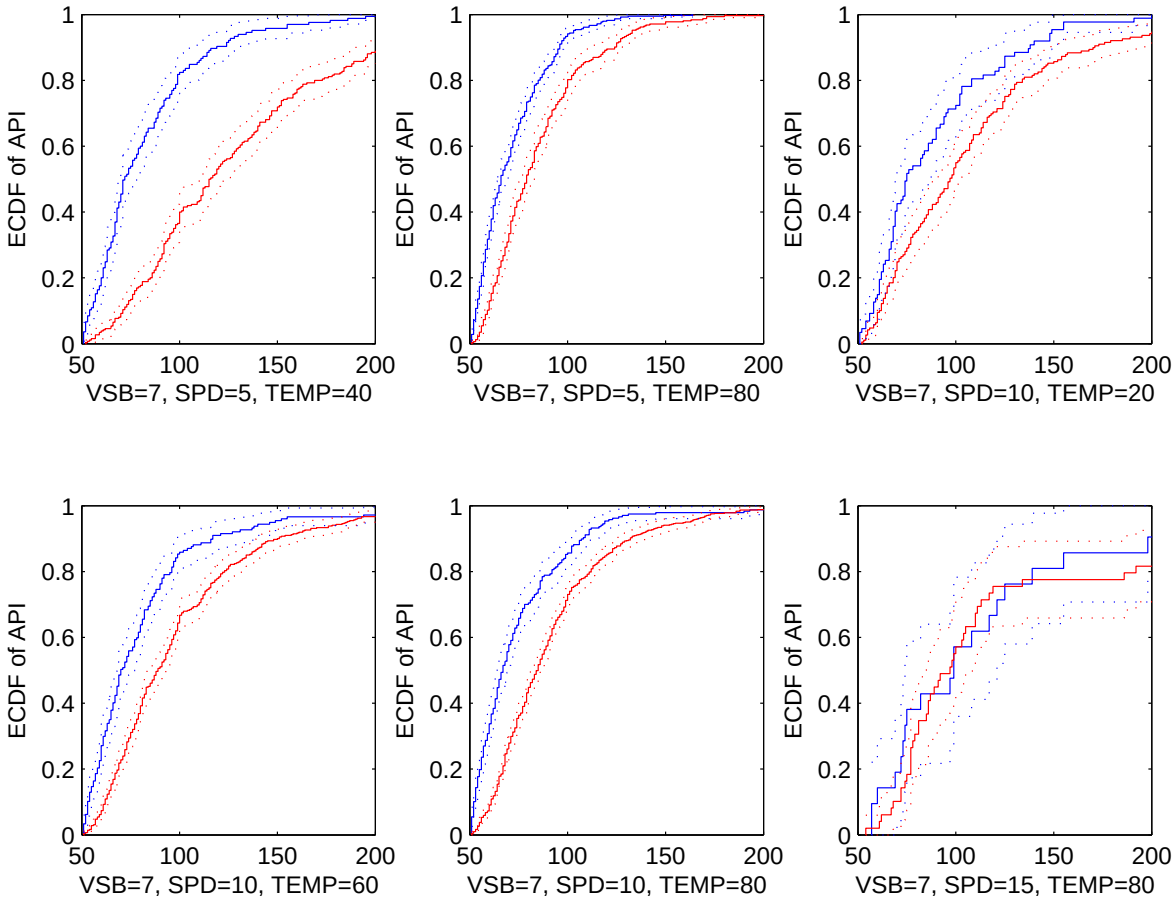


Figure SA.12: **Panel Matching Approach: Huhehaote-Baotou.** The red lines represent Huhehaote, the blue Baotou. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.

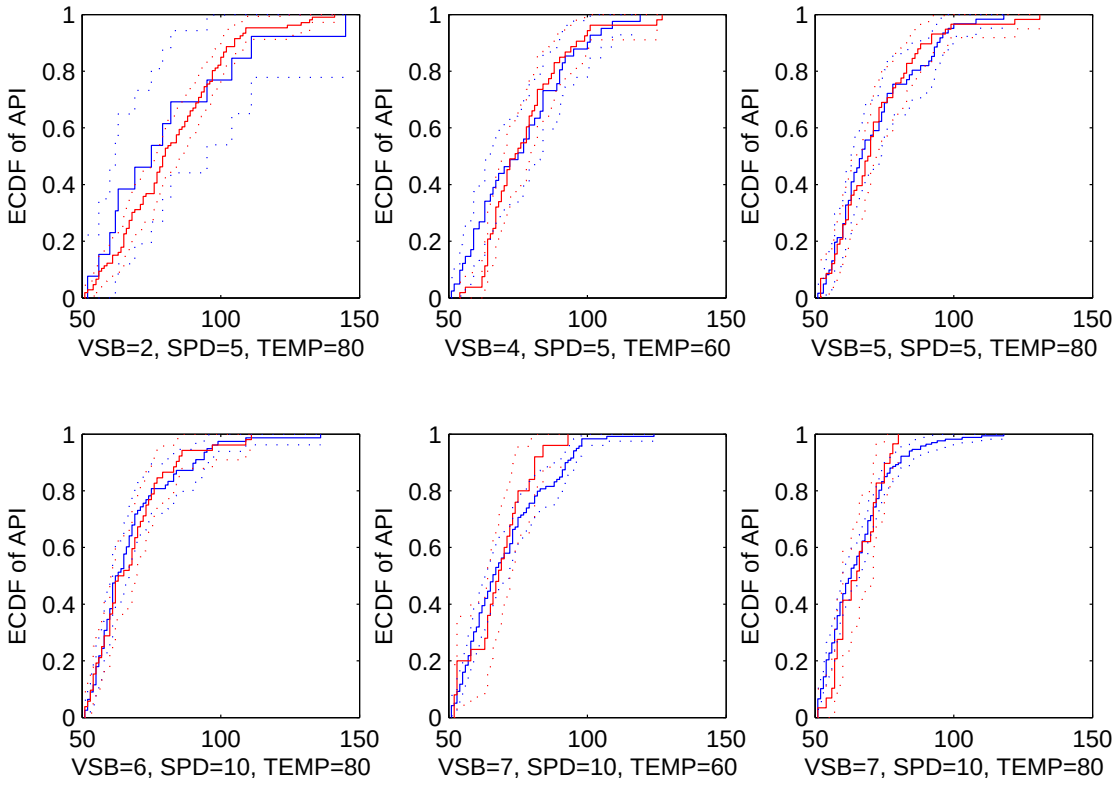


Figure SA.13: **Panel Matching Approach: Wuhu-Maanshan.** The red lines represent Wuhu, the blue Maanshan. The solid lines are the empirical cdfs of API conditional on the weather variables specified in the plot, the dotted lines give 95% point-wise confidence intervals for the empirical cdf. VSB denotes visibility, WSP wind speed, and TEMP temperature. The values for the weather variables are the discretized versions of the actual data for details on the discretization see Appendix C.