Air Pollution as a Cause of Sleeplessness: Social Media Evidence from a Panel of Chinese Cities*

Anthony Heyes

Department of Economics

University of Ottawa

Anthony.Heyes@uottawa.ca

120 University Private, Ottawa, ON, Canada, K1N 6N5

Mingying Zhu

Department of Economics

University of Ottawa

mzhu089@uottawa.ca

120 University Private, Ottawa, ON, Canada, K1N 6N5

*Heyes is also part-time Professor of Economics at the University of Sussex and a Tier 1 Canada Research Chair (CRC) in Environmental Economics. The authors acknowledge financial support for this project from the CRC and from SSHRC under Insight Grant project #435-2017-1069 "Air Pollution and Human Well-being".

Air Pollution as a Cause of Sleeplessness: Social Media Evidence from a Panel of Chinese Cities

Abstract

We provide first evidence of a link from daily air pollution exposure to sleep loss in a panel of Chinese cities. We develop a social media-based, city-level metric for sleeplessness, and bolster causal claims by instrumenting for pollution with plausibly exogenous variations in wind pattern. Effect sizes are substantial and robust. In our preferred specification a one standard deviation increase in AQI causes an 11% increase in sleeplessness. The results sustain qualitatively under OLS estimation but are attenuated. The analysis provides a previously unaccounted for benefit of more stringent air quality regulation. It also offers a candidate mechanism in support of recent research that links daily air quality to diminished workplace productivity, cognitive performance, school absence, traffic accidents and other detrimental outcomes.

Keywords: Air pollution - social costs - IV methods

1 Introduction

Our objective in this paper is to investigate a possible causal effect of urban air pollution on the sleep of city inhabitants. Air quality - particularly in cities - is one of the great policy challenges of our time. Understanding the full range of negative impacts of pollution is an essential prerequisite for welfare evaluation of policy interventions.

Sleep is an essential input to human well-being. Loss of sleep reduces mental function along various dimensions such as learning (Huber et al., 2004), memory (Diekelmann and Born, 2010), judgement (Killgore et al., 2006), speed of reflex (Maquet, 2001) and emotional balance (Ireland and Culpin, 2006). It is correlated with lower self-reported well-being (Hamilton et al., 2007; Steptoe et al., 2008). Tiredness - the inevitable consequence of sleeplessness - has been causally linked to various negative outcomes including road traffic accidents (Valent et al., 2010), workplace productivity (Zammit et al., 2010; Rosekind et al., 2010), industrial injuries (Barnes and Wagner, 2009), absenteeism (Daley et al., 2009), relationship quality (Gordon and Chen, 2014), domestic violence (Meijer et al., 2010), and compromised school performance (Chung and Cheung, 2008). In terms of health, shortage of sleep over various time scales has been linked to reduced functioning of immune systems and subsequent increased susceptibility to disease, increased risk of hypertension, cardiac and breathing problems, increased adiposity, and negative mental health outcomes.[1]

It is not a surprise that both individuals and governments invest in protecting sleep, and that individuals when asked express a substantial willingness-to-pay to avoid sleep loss (Pollinger, 2014; Delfino et al., 2008).[2] In summary, given that the typical adult in most societies spends between 7 and 8 hours of each day engaged in the activity of sleep (and children longer): "If sleep does *not* serve an absolutely vital function, then it is the biggest

[1] There is a large literature on the health implications of both short-term and chronic sleep loss (Altevogt and Colten, 2006; Cappuccio et al., 2010).

[2] For example, individuals spend on good mattresses and other aids to healthful sleep, worry about the noise environment when they buy a home,etc.. Governments spend on sleep research, impose regulation of night-time noise around airports, etc.. Employers are also aware of the benefits of sleep. See for example the lead article *Why Companies are Willing to Pay to Make Sure You Get a Good Night's Sleep* in Executive Style Magazine (21 April 2016) on the productivity benefits of well-rested employees.

mistake the evolutionary process has ever made." (Rechtschaffen, 1971).

Despite the centrality of sleep to humans, and the diverse contributions that it makes to individual and societal well-being, economic analysis of it has been cursory. Biddle and Hamermesh (1990) treat sleep choice as a time allocation problem. Similarly, Asgeirsdottir and Zoega (2011) provide a model of sleep behavior as an investment that an individual makes in the level of alertness he then enjoys during the day, in the spirit of the approach taken to health as human capital.

While the channels that might link pollution exposure to lower quantity or quality of sleep are obvious (shortness of breath, elevated heart-rate, irritation of upper airways, eyes etc.), research linking pollution exposure to sleep outcomes comprises (to the best of our knowledge) three papers. (1) Strøm-Tejsen et al. (2016) manipulate indoor air quality in the campus bedrooms of 16 students and find that indoor air quality impacts both sleep quality (as measured by subject-worn actigraphs) and next-day performance on math and language tests. (2) Using measures of outdoor air quality and subjects with sleep disorders, Zanobetti et al. (2010) show that the same-night AQI in the city in which the patient resides correlates with likelihood of episodes of sleep apnea (pauses in breathing during sleep). While this study is suggestive, the focus on those with sleep-illnesses, and the observation of subjects via a polysomnograph (sensors at nose, fingers, face and scalp) make drawing implications about a wider pollution-sleep link difficult. (3) Focusing on long-term exposure, and without using tools that would allow for causal inference, Billings et al. (2017) find a negative association between sleep efficiency amongst a sample of older people and 5-year and 1-year measures of $PM_{2.5}$ in the neighborhoods in the six US cities in which they live.

Sleep loss is a significant problem in China (Luo et al., 2013) and elsewhere. For the 10 largest Chinese cities we construct a nightly, population-level measure of sleeplessness using frequency of use of the Chinese characters meaning 'can't sleep', 'sleepless' etc. on the very widely-used social media site Weibo.[3] We use OLS to characterize a positive association

[3]We will be careful to qualify our use of the term "population level" in the data section. Population-level behavior on various internet platforms is increasingly being exploited by social scientists. Choi and Varian

between that measure and same-day local air quality. To reinforce our causal interpretation of this relationship we apply IV methods, using plausible exogenous variations in short-term wind patterns to instrument for air quality. In our preferred specifications we find that a one standard deviation of $PM_{2.5}$ increases 10.72% relative to mean. The statistical significance and effect size prove remarkably robust to a battery of alternative specifications and tests.

We are cautious not to over-interpret the results and monetizing the sleep loss caused by diminished air quality is beyond the scope of this paper, though it is worth noting that previous research does provide WTP estimates that could be exploited in a back-of-the envelope exercise. The results are instructive in two ways. First, the loss of sleep plausibly impacts the well-being of the affected individual him or herself through a variety of channels. Second, as noted, the results provide a mechanism consistent with recent research linking short-term variations in air quality to reduced workplace productivity (Zivin and Neidell, 2012; Chang et al., 2016), school absence (Currie et al., 2009), exam performance Mendell and Heath (2005), motor vehicle accidents (Sager, 2016) etc..

Section 2 details data sources. Section 3 describes methods. Section 4 and Section 5 present main and robustness results. Section 6 summarizes the results from joint estimation. Section 7 concludes.

(2012) show that Google search data can be used to predict demand for automobiles, home sales and travel behavior. Several papers demonstrate the efficacy of using internet search metrics to predict health outcomes - especially flu - and Google itself established the Google Flu Trends tool in 2008. Goel et al. (2010) show that searches can predict the success of movies, songs and video games. In an environmental application, Herrnstadt and Muehlegger (2014) show that searches for "climate change" and "global warming" in a particular US city are sensitive to short-term deviations of weather from normal. Much recent work has been devoted to Twitter-driven predictive analytics. For three examples among many: Bollen et al. (2011) show that Twitter mood can be used to add explanatory power to stock market forecasts, Gerber (2014) uses Twitter key words to predict crime patterns, and Gayo-Avello et al. (2011) are among several papers using Twitter to predict elections. A central way in which our methods depart from this literature is that we will use measures from social media as dependent variable. In that regard the paper relates to Baylis (2015) who shows the effect on unusual temperature on Twitter-sentiment.

2 Data

We investigate the effect of air pollution on sleep in the 10 largest Chinese cities. To do this we develop a nightly, city-level measure of sleep quality derived from posts on social media and connect it to high frequency data on air quality. Detailed meteorological data both to control for the likely confounding influences of weather on sleep and for the construction of our instrument.

2.1 Sleep

A challenge in this research is to develop a defensible measure of sleeplessness, that is a nightly index for how badly (or well) the inhabitants of a particular city are sleeping.

A number of surveys have asked questions about sleep.[4] However none of these provide the temporal granularity that we require (the exact date of interview and some question about short-term, ideally daily, sleep experience). Even if such questions were asked, the resulting responses would be threatened by imperfect recall of respondents, and other shortcomings typical of retrospective survey-derived data.

We exploit what people are saying on the Chinese micro-blogging Weibo. Weibo was launched in August 2009 and growth in its use was explosive, not least because most of the key social media platforms familiar to those living elsewhere (including Twitter, Facebook, Instagram and Youtube) are blocked in China. It is the biggest social media site in China, and by 2016 it had more than 503 million registered and 313 million regular users from amongst the 720 million internet users in that country (DeLuca et al., 2016). As with Twitter, messages were - at least during the period that we analyze - subject to a tight word limit. In comparison to Twitter it has a greater personal than professional orientation in the way it is used (Sullivan, 2012), with substantially more posts outside standard office hours (Gao et al., 2012). Users typically post what they see, hear and think at any moment and, while it needs to be mined with caution, the content of posts provides the researcher with a

[4] For example Chen et al. (2004), Yu et al. (2007) and Sun et al. (2015).

potential 'window' into the mind of users and a rich data source.

2.1.1 Keywords

Written Chinese is not alphabetic but rather comprises self-standing characters or glyphs. It is logo-syllabic, which means that a character represents a whole word (physical object, concept, *etc.*). Literacy requires the memorization of a large number of such characters and a well-educated Chinese person knows more than 4000, while between 2000 and 3000 are needed to read a newspaper (Norman, 1988). This is helpful to us. By its nature there are many fewer duplicative ways to express concepts than is common in alphabetic languages, such as English. "Shimian" and "Shuibuzhao" are the two characters that have meaning equivalent to that covered by English words and expressions such as "sleepless", "can't fall asleep", "losing sleep", "insomnia", *etc.*. A further advantage of Chinese is that these are used in the affirmative, so we avoid complications arising from conventions for negation that would arise in most other languages.

We search for the hourly use of these keywords in Weibo posts from users located within each of the 10 most populous cities in China (these are Beijing, Changsha, Chongqing, Guangzhou, Hangzhou, Nanjing, Shanghai, Wuhan, Tianjin and Zhengzhou). Weibo offers advanced search tools that enable users to obtain all public posts filtered by keyword, date, time period (minimum duration 1 hour), and location (city). In contrast to Twitter - which limits the number of tweets that can be searched to the 1% in the Streaming API - Weibo allows for search of the entire corpus of posts.[5] We use these to construct a panel of the number of posts featuring the keywords of interest for each hour of each night (11pm through 7am), for each city for the two year period 2014 and 2015.

It is worth reflecting on this as a dependent variable. The question is not whether keyword use on Weibo is a perfect measure of the thing that we want to measure (the extent to which

[5]However, it only presents the first 1000 results from any search. Though potentially constraining on our data collection exercise, that our searches are within-city, within-hour, means that in practice in no case does this limit bind.

inhabitants of a particular city are sleeping on a particular night) - of course it isn't. Rather, is it a good enough measure, and is it better than others available?

There are two main challenges to our claim that intensity of use of the words "shimian" and "shuibuzhao" provides a valid proxy for city-level sleeplessness.

First, perhaps other terms exist that might be used to express the difficulty sleeping that we fail to consider. Inspection of Chinese thesauri and discussion with Chinese speakers make us doubt that this is the case. However, even if it were it is unlikely to disturb our conclusions. (1) The correlation between use of "shimian" and "shuibuzhao" in our sample is very high (0.96) and the ratio between use of one and use of the other proves insensitive to air quality conditions. We use the word counts as an index, rather than focus on absolute levels. If an additional synonym exists that we have ignored, then provided its use is closely correlated with these two then its exclusion is not a concern.[6] (2) Measurement error in the dependent variable that such an oversight would imply does not bias OLS or 2SLS estimates, only reduces their efficiency. We also investigate and refute the possibility that what we are picking up is a simple a proxy for overall Weibo use by showing that the sleep metric is uncorrelated with the use of a series of sleep-neutral words (table, cat, etc.), with appearances of the latter not systematically sensitive to air pollution conditions.

Second, Weibo users are not representative of the Chinese population in general. In particular users are younger, more educated and higher income than the broader population (Chan et al., 2012; Chiu al., 2012). While results should most properly be seen as reflecting a treatment effect in the Weibo-using part of the community we do not see this as problematic. These are likely the high value workers in Chinese urban society and disturbance of their sleep can be expected to have correspondingly important economic impact. Further, there is no reason to think that effects observed in this group would not be observed in the non-Weibo-using part of the population. Indeed, it is plausible to think that those effects could be larger for at least two reasons: (1) In terms of self-protection, those with internet access

[6] A problem would arise for us if there was an excluded means of expression whose comparative intensity of use varied systematically with air quality conditions. This seems implausible.

are disproportionately likely to own both air conditioners and air purifiers. (2) Weibo-users are younger than the general population, and most physical effects of pollution are more pronounced among the old. However this is a useful caveat to carry in mind.

2.2 Pollution

Data on pollution at our locations of interest was collected from www.aqistudy.cn. This website compiles real-time data on pollutants from the Chinese Ministry of Environmental Protection (MEP) and converts it into daily average measures. The pollutants for which we have data are $PM_{2.5}$, CO, NO_2, and O_3 (in addition to AQI).[7] Summary statistics for daily ambient measures in our whole sample are included in Table 1 (and by city in the Appendix Table A1).

Table 2 defines the categories of air quality days as defined by the Chinese government for each pollutant and - in the right hand column - the percentage of days in our sample that fall within each category on the AQI measure.

Table 3 summarizes the correlation between daily city-level measures of the individual pollutants in our sample. In a number of cases the correlations are quite high, often exceeding 0.6. Most of our analysis will be conducted pollutant by pollutant, only later including all pollutants in the same regressions. This follows Schlenker and Walker (2016).

Our analysis is conducted at city-level and we calculate air quality measures by taking a simple arithmetic mean of data from all monitors within a city (the number of monitors within our 10 cities varies between 9 and 17). While we know that a user is based in a particular city we do not know precisely where, nor his or her movements during the day. To allay concerns about intra-city variations in pollution conditions we calculate the correlations between readings at each pair of monitors in each city. The results are reported in Appendix Table A2 (and for illustration in detail for Beijing in Appendix Table A3 through A7).

[7] Historically the quality of official data on air quality in China has been questioned. In particular there has been evidence of manipulation around key thresholds (Chen et al., 2012). Stoerk (2016) tests the consistency of official data with Benford's Law, and with US Embassy data, and concludes that it is reliable from 2013.

With the exception of CO - a more localized pollutant - pairwise correlations are very high, typically close to or above 0.9. In other words pollution measured at any particular monitor is a good indicator of levels across the city.[8]

2.3 Weather

Disentangling the potentially confounding effects of weather is important. Weather conditions (in particular temperature, humidity, precipitation) can influence sleep directly (Okamoto-Mizuno and Mizuno, 2012; Van, 2006).

Meteorological data are obtained from the weather stations registered by the World Meteorological Organization (WMO) that are collated by the National Oceanic and Atmospheric Administration (NOAA). The weather variables comprise average temperature ($°C$), maximum temperature ($°C$), minimum temperature ($°C$), average humidity (%), maximum humidity (%), minimum humidity (%), sea-level pressure (Hpa), wind speed (Km/h), wind direction (°) and precipitation (mm). We combine the hourly weather data into daily mean levels corresponding to the daily average air pollution levels of each city. Summary statistics for the dataset appear in Table 1 (and for each city separately in the Appendix Table A1).

3 Methods

We investigate a link from air pollution in city i on day t to our city-level metric for sleeplessness in that city on that night. In simple terms: If the air in Nanjing is highly polluted today, does that damage the quality of sleep in Nanjing tonight?

[8]Insofar as measurement error exists in this regressor we expect it to attenuate OLS estimates, implying that the effect sizes identified under OLS should be interpreted as *under*-stating true effects. The coefficients from the IV exercise will not be subject to such bias.

3.1 OLS

We first use OLS to estimate the association between air quality and sleeplessness in a straight-forward panel fixed effects setting. We estimate the following specification

$$lnS_{it} = \alpha_0 + P_{it}\beta + W_{it}\gamma + \theta_i + \lambda_t + \epsilon_{it}. \tag{1}$$

S_{it} is the sleeplessness index in city i on the night following calendar date t. lnS_{it} denotes that the outcome variable is logged.

P_{it} is the daily average pollutant concentration in city i on date t. The primary pollutants that we consider in turn are $PM_{2.5}$ and the composite AQI measure.

We control for a wide set of potential confounders. W_{it} is a vector of weather controls containing average temperature, maximum and minimum temperature, humidity, precipitation, average wind speed, sea-level pressure. The temperature and humidity measures enter as indicators or 'bins' (5 $°C$ indicators for average temperature, 20 % indicators for average humidity) to accommodate possible non-linear effects.[9] θ_i is a city fixed effect that controls for time-invariant city characteristics. λ_t is a vector of time fixed effects, comprising year_season, day of week and a dummy for holiday dates. ϵ_{it} is the error term.

Our coefficient of interest is β, which relates air pollution to sleeplessness. It can be interpreted as 100*β% increase in sleeplessness due to additional unit of pollutant. Most of the effect sizes that we will report are based on the percentage change due to one standard deviation of pollutant, which could be computed by multiplying 100*β by one standard deviation (47.843 for $PM_{2.5}$ and 55.557 for AQI).

3.2 Single pollutant versus joint estimation

Our initial results will be derived from single pollutant models in which regressions are run that incorporate PM2.5 without co-emission. There is also an AQI variant, where AQI

[9]Results prove similar under quadratic estimation, a popular alternative approach to non-linearity.

is a composite measure that captures the 'binding' pollutant on any particular date. We report the joint estimation exercise in Section 6.

Note that research in this area is plagued by the difficulties of disentangling the effects of *particular* pollutants from the overall cocktail of pollutants that an individual will typically be inhaling on a 'bad air' day.

Some settings allow for a clean route around this problem. A nice recent example is Lavaine and Neidell (2017). Helpfully for them the oil refinery strikes that they exploit as exogenous events that temporarily improved air quality in a set of French towns acted on sulphur dioxide in particular, leaving ambient levels of other key pollutants undisturbed. But often the inclusion or exclusion of pollutants is driven by data availability in particular settings. Papers typically report results of regressions that include a single (or limited subset) of pollutants. For example, among well-known investigations of the effect of short term air quality variations on various outcomes; (1) Zivin and Neidell (2012), who look at productivity of agricultural works, appoint ozone as their pollutant of interest and control only for $PM_{2.5}$. (2) Ransom and Pope (1992), looking at school absences, exploit data only on PM_{10}, finding negative effects.[10] (3) Ebenstein, Lavy and Roth (2016), studying the effect of daily pollution levels on the exam performance of Israeli children, consider only $PM_{2.5}$.[11] (4) Schlenker and Walker (2016), looking at the health impacts of pollution, deploy data on only CO, NO_2 and ozone, and their main results are derived from specifications in which each pollutant is used as explanatory variable sequentially, without controls for the other two (indeed all but one of the eight tables in Schlenker and Walker (2016) report results of single pollutant exercises). They later insert the three pollutants in the same regression with qualitative loss of results.[12]

[10] In the pursuant literature various authors have considered varying permutations of the major pollutants. For example, Gilliland et al. (2001) add ozone and NO_2 and find *beneficial* effects of PM_{10} on absences. Currie et al. (2009) study three of the main pollutants, CO, PM_{10} and ozone.

[11] While in an earlier version (Ebenstein et al., 2016) they also investigate CO, they did not do so simultaneously, and were unable to account for other major pollutants.

[12] They are explicit in "...acknowledging that we may be picking up the health effects of other pollutants" (page 787). The omission of $PM_{2.5}$ and PM_{10} - with clear links to a variety of cardiovascular and other health outcomes - is a challenge for the interpretation of their results. In an Appendix exercise they note

We are to some extent insulated from these problems because our main estimates derive from IV methods. However, given the (sometimes strong) covariance between pollutants we will follow Schlenker and Walker's caution in tying effects to particular individual pollutants. As it turns out our results will all work in the same direction - more pollution causing greater sleeplessness. But we are more confident interpreting this as a story about 'dirty air', and circumspect in pollutant by pollutant inference.

3.3 IV

There are several challenges to the validity of OLS estimation here.

First, likely measurement error in pollution. Our theoretical foundation is predicated on the possibility, founded on plausible physiological foundations, that exposure of an individual to elevated levels of pollution increases the chance of disturbed sleep. However, we observe ambient air quality (which we have shown to be comparatively uniform across monitor sites within a particular city on a particular date) rather than individual exposure. For example, we do not observe self-defensive behavior, such as closing of windows and use of air purifiers, which can reduce effective exposure.[13] The measurement error this would imply in the dependent variable would lead to attenuated OLS estimates of our coefficient of interest.[14] Second, while we included a rich set of controls for potential confounders - taking particular care with weather - we cannot rule out omitted variables. For example, air pollution may be positively correlated with unobserved variations in city-level economic activity, which may in turn influence sleeplessness through other channels.

For these reasons we supplement our OLS analysis using two-stage least squares (2SLS),

that this is down to absence of data. As such they conclude that: "We believe that some amount of caution is warranted in interpreting CO as the unique pollution-related causal channel leading to adverse health outcomes; there may be in fact other unobserved sources of ambient air pollution that covary with CO that may also effect health" (page 800).

[13] In some sense this doesn't matter. What we end up with is not an individual level sleep 'production function' but a population-level effect from ambient conditions to sleep. In terms of defensive behaviors, our results should be interpreted as incorporating such margins of adjustment.

[14] In their investigation of the effects of short-term exposure to health, Moretti and Neidell (2011) provide evidence and insightful discussion of the problems associated with measurement error in this context.

with an instrument based on wind direction.

3.4 Instrument

Air pollution in Chinese cities is known to be highly sensitive to wind direction and speed, as pollutants are carried from neighboring cities (Fu et al., 2017). Ambient pollutants, especially fine particles can travel over a long distance by wind, ranging from hundreds to thousands of kilometers (EPA, 1996). The fact that airborne particles can be transported by wind and affect the places on the downwind side has been used in linking air pollution to health outcomes. For example by Schlenker and Walker (2016) in their study of adverse health effects downwind of airports. Bayer et al. (2009) use pollution levels in nearby (but further than 80km) cities to instrument for local pollutant levels. There are also studies that focus on estimating movement of air pollutants between cities (for example Chen and Ye (2015)). We develop an instrument based on plausibly exogenous day to day variations in wind patterns which, consistent with the existing literature, proves to have strong relevance (delivers a strong first stage). The method is similar to that applied by Schlenker and Walker (2016), but whereas they exploited a single source of supply of pollution (an airport) to any particular neighborhood, our study cities typically import wind-borne from multiple neighboring cities, requiring that we apply an intuitive weighting scheme.

For each study or target city i - recall we consider the ten most populous in China - we identify other smaller cities located (centre to centre) within between 100km and 200 km. These are likely sources of pollution imported to city i if the wind happens to blow in the 'right' direction. We refer to these as 'source' cities for city i. Neighboring cities within 100km are excluded to minimize risk of endogeneity (Bayer et al., 2009; Zheng et al., 2014).[15] Source cities and their coordinates are listed in Appendix Table A8.

We deliberately take a 'standard' approach to constructing our first stage, which is

[15] Bayer et al. (2009) exclude the distant sources within 80km, Zheng et al. (2014) within 120km. In their study of medium term health effects of $PM_{2.5}$ and SO_2, Barreca, Neidell and Sanders (2017) allow for the transport of pollution from a single power station up to 100 miles (161 km). In a robustness check we consider the effect of varying these cut-off distances and find results largely undisturbed.

$$P_{it} = \eta_0 + \psi P_{source_{it}} + W_{it}\gamma + \theta_i + \lambda_t + \epsilon_{it} \tag{2}$$

where

$$P_{source_{it}} = \sum_{j}^{J} \omega_{ijt} \overline{P_{jtmonth}}$$

P_{it} is actual pollution in target city i on date t. The coefficient of interest is ψ and captures the effect of pollution from upwind source cities on the target city.

$P_{source_{it}}$ is an index that proxies the amount of pollution expected imported into target city i from source cities on a particular day. It is important that the construction of this index is fully-understood so we will describe its components in some detail. Validity of the instrument will require that the only way in which wind directions influence sleep patterns in the target city is through induced changes in target city air quality.

$\overline{P_{jtmonth}}$ is the mean level of pollution in source city j in the associated month. In other words a measure of how 'potent' a particular source is as a supplier of pollution. As is well known, transport of pollution from source to target city on a particular day depends upon wind direction and speed. In particular, other things equal imports of pollution from city j on air to city i are greater when; (a) the city is close, (b) windspeed is high on a particular day, (c) the angle between wind direction and an imaginary line joining the two cities is small (Zahran et al., 2017; Anderson, 2015; Schlenker and Walker, 2016). The vector of weights ω_{ijt} capture this. In particular we inverse-distance weight the source cities (Equation 3) where geographical distance is adjusted to allow for windspeed and angle (Equation 4).

$$\omega_{ijt} = \frac{\frac{1}{trans_{jt}}}{\sum_j^J \frac{1}{trans_{jt}}} = \sum_{j}^{J} \frac{\frac{1}{trans_{jt}}}{\frac{1}{trans_{1t}} + \frac{1}{trans_{2t}} + \frac{1}{trans_{3t}} + ... + \frac{1}{trans_{Jt}}}, \tag{3}$$

where

$$trans_j = \frac{dj}{windspeed_i * \cos|\phi_i - \phi_j|_{>0}} \tag{4}$$

Wind direction can vary during the course of a day. We use daily average direction constructed from hourly data, consistent with first principles and most existing studies (including Schlenker and Walker (2016) and Herrnstadt et al. (2016)). Only positive values of $\cos|\phi_i - \phi_j|$ are included when the index is calculated, *i.e.* attention is limited to source cities that are (not necessarily directly) downwind on any particular day.[16] This occurs where the difference between wind direction and the direction of the vector between cities j and i is less than 90 degrees. In a robustness check we find that results are largely undisturbed if we instead limit to those where the difference is no greater than 60 degrees. The complexities of pollution transport by wind do not allow us to specify fully the process whereby pollution from one city influences air quality in another, but the functional form here is a simplified version standard in modelling of this sort. For a recent application, the analysis here coincides with Schlenker and Walker (2016) who account for the cosine of variation of wind direction from point source (airport) to centre point of zipcode. Importantly, it is unlikely that the precise functional form adopted here would influence the defensibility of the exclusion restriction. Moreover, we will try some alternatives for the purposes of robustness later. Relevance of the instrument is assessed statistically at the first stage.

[16]The angle between wind direction and the line joining the central points of cities i and j is $|\phi_i - \phi_j|$. All angles are measured in degrees clockwise from due North (0° and 360° equal North). The cosine transformation implies a particular weighting to sources at different angles. Recall that the cosine of zero degrees is 1, cosine of 20 degrees is 0.93, cosine of 60 degrees is 0.5 and so on. So other things equal a source 60 degrees off the wind line carries half the weight as a source that is directly upwind. The weighting is consistent with first principles (Anderson, 2015). Later we show that the results are qualitatively robust to dropping the weighting scheme altogether. As would be expected the precision of estimates is compromised, though significance of results is maintained.

3.5 Lagged IV

As noted, in our base specifications we limit attention to source cities located 100 to 200 km from the target city ($100km < d_{ij} < 200km$). Airborne pollutants leaving one city take more time to transport over a greater distance, which points to a delayed impact on the target. Our primary measure of pollution is average ambient concentrations from midnight to midnight, and the outcome of interest is sleeplessness in pursuant night (11 pm to 7 am). With average wind speed in the sample at around 8 km/h transport of air from a city at distance of 100 km would take over 12 hours, from 200km 24 hours. To capture this some specifications include a one day lag,

$$P_{it} = \eta_0 + \psi_{it-1} \sum_{j}^{J} \omega_{ij(t-1)} \overline{P_{j(t-1)month}} + \psi_{it} \sum_{j}^{J} \omega_{ijt} \overline{P_{jtmonth}} + W_{it}\gamma + \theta_i + \lambda_t + \epsilon_{it} \qquad (5)$$

We expect each of the coefficients ψ_{it-1} and ψ_{it} to be positive and similar in order of magnitude. In unreported analysis we have tried alternative specifications with additional lags without disturbing results discernibly.

4 Results

4.1 OLS

Table 4 reports the coefficients from estimating equation (1) using OLS regression model for AQI (Panel A) and $PM_{2.5}$ (Panel B), where the dependent variable is log form of sleeplessness and the independent variable of interest is daily pollution.

Each of the 10 coefficients reported in Table 4 is derived from a separate regression. We will talk for now about coefficient sizes, and return to interpret the effect size that they imply later.

Column (1) is the sparsest specification and includes only city fixed effects, netting out

any unobserved, time-invariant city characteristics (size, Weibo-penetration, building characteristics, etc.). Reading down this column we see positive coefficients for each pollutant, in each case significant stronger than 1%.

From Column (2) to Column (4), we add time controls (year_season, day of week and holiday fixed effects one by one). As expected seasonal effects have an important impact on sleep. Sleep behavior can be expected to be different on weekdays to weekends, and holidays to non-holidays. The inclusion of these has little effect on the estimated coefficients on the pollution regressors in Column (2).

In Column (5) we allow for weather effects. The weather controls include bins for average temperature and humidity, and linear measures for precipitation, sea-level pressure, wind speed, maximum and minimum temperature and humidity. Weather effects are known have a meaningful impact both on sleep (not presented in this table) but, more importantly for us, the strength of the relationship between air quality and sleeplessness. The coefficient increases a bit compared with that under temporal controls.

Standard errors are clustered at city level. As there are only 10 clusters (cities) we use the pairs cluster bootstrap method (Cameron et al., 2008; Harden, 2011), one of the most versatile remedies for small numbers of clusters. The likely alternative approaches would have been cluster-adjustment of the t-statistics (Bakirov and Székely, 2006) and wild cluster bootstrap (Cameron et al., 2008). In a robustness check we will verify that these alternatives would not have disturbed inference.

For the OLS part of the paper, Column (5) summarizes the preferred specification.

While the sign and significance obtained for coefficients on all pollutants in this section provides valuable insight, earlier we identified concerns - in particular measurement error related to effective pollution exposure levels - that led us to expect attenuation in estimated coefficient values. Insofar as these concerns are valid we would expect the effects summarized in the last paragraph to *under*-state true effect sizes. To address this concern we now report IV estimates.

4.2 IV

The main IV results are reported in Table 5. From Columns (1) to (6), city fixed effects, temporal controls and weather covariates are added in sequence. Each column reports the outcome of a separate regression, and for $PM_{2.5}$ and AQI we run alternatives without and with the lagged instrument included in the first stage (odd and even numbered columns respectively). All the regressions include the full suite of controls already described.

The dependent variable in the first stage is daily-mean pollutant in target city i and the IV is the weighted average pollution of surrounding source cities. Recall that cities are included if they are between 100km and 200km in the upwind direction, where upwind is defined as within 90 degrees of the average within-day wind direction.

The first stage exercises work well. We find a strong effect of variations in pollution in source (upwind) cities on the target city. In each case significance is achieved at better than 1%. The lagged pollutant measure is also significant in both cases, as anticipated. The F-statistics in each of the eight first-stages are very high. So we have no concerns about weak instruments.

The second stage replicates the preferred OLS specification, regressing the daily sleeplessness measure on the predicted level of pollution obtained from the first stage.

In each case the coefficients on instrumented pollution are very similar between odd and even rows. Though the lagged pollution measure 'matters' in the first stage, its inclusion has relatively little impact on the coefficient of interest in the second.

Our preferred specifications from Table 5 is Column (6).[17] In each case the estimated coefficients are times larger in absolute size than those derived from OLS, consistent with our expectation that the estimates from the latter were attenuated.

A one standard deviation increase in $PM_{2.5}$ causes an increase in sleeplessness equal to 10.72% of the daily mean. For AQI a one standard deviation increase causes similar

[17] In addition, and following Schlenker and Walker (2016) Table 1, we explore the possibility that pollution may be dispersed by high winds by adding an interaction term $P_{source_{it}} * windspeed_{it}$ to our preferred first-stage specification. This has little impact on results - summarized in the Appendix Table A9.

sleeplessness, amount to 10.89% of mean level.

5 Robustness and Falsification

5.1 Selection of source cities

In developing the instrument two decisions were made as to how source cities were to be selected. First, we considered cities more than 100 but less than 200 km distant (*i.e.* $100km < d_{ij} < 200km$). Second, to be considered 'upwind' the angle between the wind line and a straight line drawn between source and target city had to be less than 90 degree (*i.e.* $|\phi_i - \phi_j| < 90°$). Since source cities are described by their monthly average pollutant characteristics, and locations do not move, the only variation in source city across dates comes from plausibly exogenous day to day variations in wind direction. Here we conduct two robustness tests on these thresholds.

In Table 6 we report the results of re-estimating the preferred IV specification but with cities selected as sources if they lie within a narrower, 60 degree angle of the wind line (*i.e.*, $|\phi_i - \phi_j| < 60°$). The results from the first and second stages look very similar to those reported in Table 5. Sign and significance is maintained throughout and coefficient values are little disturbed.[18]

Next we restore the preferred assumption on wind angle wind to 90 degrees, but expand the thickness of the 'donut' from which source cities are drawn, in particular selecting source cities at a distance $100km < d_{ij} < 300km$ (rather than $100km < d_{ij} < 200km$). The results of this exercise are summarized in Appendix Table A10. For both AQI and $PM_{2.5}$ the results are similar to those in Table 5.

[18]The F-statistics from the first stages are somewhat smaller, though still good. This reflects that building the instrument on a basis that excludes source cities at $60 < |\phi_i - \phi_j| < 90$ means that we lose part of correlation of the instrument with target city pollution.

5.2 Reduced form and drop weighting scheme

Table 7 reports the results of a reduced form. Columns (1) and (2) reproduce the OLS and IV results respectively (that is the coefficients in Column (1) coincide with those from the OLS regressions in Table 4. Column (2) repeats the second stage results under Columns (5) in Table 5. Column (3) re-generates the IV results but discarding the weighting procedure. In other words we replace $P_{source_{it}}$ with a simple arithmetic mean of same-day pollution levels in upwind source cities (in those cities for which $100km < d_{ij} < 200km$, $|\phi_i - \phi_j| < 90°$ on a particular date). The results are very similar with those in Column (2) in which the instrument is built by weighted average of upwind pollutants.

Column (4) reports the reduced form exercise in which $P_{source_{it}}$ is the regressor of interest in an OLS regression with lnS_{it} as the dependent variable. In other words from:

$$lnS_{it} = \alpha_0^{'} + P_{source_{it}}\beta_{up} + W_{it}\gamma^{'} + \theta_i + \lambda_t + \epsilon_{it}^{'} \tag{6}$$

Again, each coefficient in this table comes from a different regression. As expected, the estimates from the upwind variant remain significant - the usual reduced form 'works' - and the effect sizes are somewhat smaller than those from the IV.

As a further test we estimate the same reduced form (Equation (6)) using arithmetic average pollutant among upwind cities. The results of this exercise are reported in Column (5). Again, as expected the coefficients from the upwind exercise are positive and statistically significant.

5.3 Precipitation

The confounding role of rain is a potentially important challenge to our inference. Introspection suggests that rainfall - either contemporaneously, or lagged through effect on mood etc. - might plausibly inhibit sleep.

While we include controls for daily-average precipitation amongst our weather controls we further probe this possibility by conducting two sub-sample exercises.

First, we re-estimate our preferred specifications on that sub-sample of days on which recorded night-time precipitation (from 11 pm to 7 am) in the target city is zero. This causes us to lose around 18% of the sample. The results of this exercise are reported in Column (2) of Table 8 and Column (2) of Table 9 for OLS and IV respectively. Results are little disturbed. This implies that the effects observed are not driven by contemporaneous rainfall.

Second, we re-estimate our preferred specifications on that sub-sample of days on which recorded precipitation during the night in question *and* the whole of the preceding calendar day in the target city is zero. This causes us to lose around 38% of the sample. The results of this exercise are reported in Column (3) of Table 8, and Column (3) of Table 9 for OLS and IV respectively. The signs and magnitudes of the coefficients are in all cases quite similar (IV estimates in each case in fact become somewhat larger than those derived from the whole sample). The level of statistical significance obtained is sustained in almost all cases - better than might have been anticipated given the considerable erosion of sample size.

5.4 Beijing and environs, Shanghai and environs

While we derive results from a panel of the 10 most populous cities in China, a further concern might be that the results are driven by a small subset.

In an unreported exercise we rerun our preferred specifications on restricted samples of cities, dropping each individually in turn, and in no case do we observe more that slight disturbance of results. However in this section we report the impact of dropping clusters of cities that may exhibit particular features that might be driving results. In particular, (1) First, we exclude the cities of Beijing and Tianjin (the Beijing-Tianjin corridor is the country's most heavily industrialized 'rust-belt' area (Shao et al., 2006); (2) Second, separately we exclude the cluster of south-eastern coastal cities of Shanghai, Hangzhou and Nanjing (these are less polluted, less industrialized, and more influenced by coastal effects).

The results of these exercises are summarized in Columns (2) and (3) of Appendix Tables A11 and A12 for OLS and IV respectively. Again, results are little-disturbed. The first stages continue to work well, and the second stage estimates are largely robust.

It is also concerned that whether air pollution remains its health effect across the ten cities in the sample. Both OLS and IV estimators of individual city are reported in Appendix Tables A13 and A14 for AQI and $PM_{2.5}$ respectively. Although the magnitude of the effects varies across the cities, most of them still have a significant impact on city sleeplessness.

5.5 Alternative standard errors

In the calculation of standard errors in the main tables we chose to bootstrap cluster at city level, judging this to be broad enough to account for the potential correlations among regressors and errors within clusters.

However, this approach delivers only ten clusters (each with 730 observations) which Angrist and Pischke (2008) suggest may be too few. Cameron et al. (2008) show that small cluster numbers can bias downwards cluster-robust standard errors leading researchers to overstate the statistical significance of results.

To investigate this threat we follow Cameron and Miller (2015) and Esarey and Menger (2015) in evaluating, for our central specification, statistical significance using two alternative approaches; (1) pairs cluster bootstrap (Bertrand et al., 2004; Cameron et al., 2008; Harden, 2011; Esarey and Menger, 2015) and, (2) wild cluster bootstrap (Cameron et al., 2008; Esarey and Menger, 2015). The results of these alternative approaches to calculating standard errors are reported in the first two columns of Tables 10 and 11 (of course coefficient estimates are unchanged across cases). As can be seen, statistical significance proves robust at conventional levels.

A separate concern related to standard errors is that spatial correlation can in some circumstances bias standard errors and so invalidate inference (Hoechle, 2007). To investigate this possibility in our setting we apply the methods of Driscoll and Kraay (1998). They

introduce a non-parametric covariance matrix estimator where standard errors are assumed heteroscedastic, auto-correlated with $MA(q)$ within panel (each city), and potentially correlated among panels. The method is appropriate for panels with small numbers of panels (in our case 10) but many observations per panel (730). The results of this exercise (for $q = 5$, though very similar results emerge with different values) are reported in Column (3) of Tables 10 and 11. Again statistical significance is maintained at conventional levels.

We also supplement the analysis with more clusters by cluster the robust standard errors at city_year_season (80 clusters), city_year_month (236 clusters) and city_year_week (1040 clusters) to check whether the significant inference changes under large number of clusters. The results are displayed in Appendix Table A15 and A16 for OLS and IV respectively. All the estimators remain their conventional significance and some even more significant.

6 Joint Estimation

Disentangling the independent effects of particular pollutants is a challenge that for research on both health and non-health outcomes. Various authors have addressed the problem in different ways, typically this involves excluding a subset of the potentially confounding substances altogether (often due to data limitations). If pollutants tend to positively covary then this leads to effects being loaded onto that pollutant or subset of pollutants that are included.[19]

Ambient levels of the various pollutants (with the exception of ozone) positively covary.[20] Some of the pollutants are precursors in the production of ozone. Furthermore the overall impact of a particular cocktail of pollutants may depend upon their mixture in complex ways. This leads us to be cautious in interpreting the results reported thus far. Taken collectively we believe that Tables 4 though 11 provide a compelling case that polluted air has a causal

[19] A different approach taken in some recent work (for example Gendron-Carrier et al. (2017)) exploits data from NASA satellites that measures Aerosol Optical Depth (AOD). AOD in effect measures how optically 'thick' the air is over a particular GIS point, but does not allow for pollutant-by-pollutant inference.

[20] In our dataset, the correlation between $PM_{2.5}$ and CO is 0.752, between $PM_{2.5}$ and NO_2 is 0.652, and between CO and NO_2 is 0.624.

impact on city-level sleep quality. While the results for $PM_{2.5}$ and CO appeared the most resilient we are wary about attributing effects to a particular pollutant too confidently.

For completeness we summarize in Table 12 the results of including all pollutants in the regressions simultaneously - the so-called 'horse race' regressions.

Columns (1) and (3) repeat the separate OLS and IV results from the single pollutant models already presented. Column (2) includes $PM_{2.5}$, CO, NO_2 and O_3 in an OLS regression. Exactly as in Schlenker and Walker (2016), signs become mixed. $PM_{2.5}$ and CO remain their positive signs but lose the significance.

Column (4) follows the method proposed by Schlenker and Walker (2016), Knittel et al. (2016) and Sager (2016). In our case, different pollutants are instrumented by their corresponding levels in source cities, and the instrumented pollution levels then included simultaneously in the same regression.[21] The coefficient on $PM_{2.5}$ remains positive and are comparable in magnitude to those from the single pollutant exercises, significance is lost a bit to 10% level.

An alternative approach - adopted by Moretti and Neidell (2011) - is to instrument for one pollutant at a time, in each case including the other pollutants as linear controls in both the first and second stage regressions.[22] In that case the coefficient on the instrumented pollutant is unbiassed, though those on the control pollutants are not. Column (5) reports the results of conducting that exercise repeatedly, with each pollutant in turn being the one that is instrumented. This approach restores significance on $PM_{2.5}$ as that in Column (4), with coefficient estimates in each case somewhat larger with double size of the conventional one.

[21] To be clear, while each coefficient in Columns (1) and (3) is derived from a separate regression, Column (2) and (4) each report a single regression.

[22] More concretely, Moretti and Neidell (2011) instrument for ozone, and include controls (uninstrumented) for CO and NO_2. They do not include measures for particulate matter.

7 Conclusions

Sleep is a central contributor to human well-being, and its disturbance has been linked to a wide set of negative outcomes. If pollution in a city has a significant detrimental impact on how the inhabitants of that city sleep, this would imply a hitherto unaccounted for social cost of air pollution. Understanding the full range of channels through which pollution effects welfare - and by implication the benefits of clean air - is a prerequisite for the design of welfare-maximising policy interventions in this area.

We provide what we believe to be the first evidence that air pollution on a particular day has a causal impact on sleep quality in a city on the following night. The effect is substantial. For the composite air quality index (AQI), notionally moving from a median clean decile day to a median dirty day (in other words from the 5th to the 95th percentile when days are ranked from clean to dirty) increases city-level sleeplessness by 33.1% of its mean value. For $PM_{2.5}$ that number is 29.7%. The estimates prove robust to a wide set of checks.

The analysis provides further evidence of the susceptibility of individual and social outcomes to anthropogenic pollution. We have argued that sleep loss is an important outcome in it's own right, but also that it can provide a mechanism to underpin a suite of less proximate outcomes identified in recent research. Further validation of the results, using alternative metrics and instruments, is planned in future research.

References

[1] **Altevogt BM, Colten HR (Eds.).** 2006. "Sleep Disorders and Sleep Deprivation: an Unmet Public Health Problem." *National Academies Press.*

[2] **Anderson ML.** 2015. "As the Wind Blows: The Effects of Long-term Exposure to Air Pollution on Mortality." *National Bureau of Economic Research.*

[3] **Andrew May.** 2016. "Why Companies are Willing to Pay to Make Sure You Get a Good Night's Sleep." *Executive Style Magazine*, Apr 21.

[4] **Angrist JD, Pischke JS.** 2008. "Mostly Harmless Econometrics: An Empiricist's Companion." *Princeton University Press.*

[5] **Asgeirsdottir TL, Zoega G.** 2011. "On the Economics of Sleeping." *Mind & Society* 10(2):149.

[6] **Bakirov NK, Székely GJ.** 2006. "Student's t-test for Gaussian Scale Mixtures." *Journal of Mathematical Sciences* 139(3):6497-505.

[7] **Barnes CM, Wagner DT.** 2009. "Changing to daylight saving time cuts into sleep and increases workplace injuries." *Journal of Applied Psychology* 94(5):1305.

[8] **Barreca AI, Neidell M, Sanders NJ.** 2017. "Long-run Pollution Exposure and Adult Mortality: Evidence from the Acid Rain Program." *National Bureau of Economic Research.*

[9] **Bayer P, Keohane N, Timmins C.** 2009. "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management* 58(1):1-4.

[10] **Baylis P.** 2015. "Temperature and Temperament: Evidence from a Billion Tweets." Energy Institute at HAAS Working Paper.

[11] **Bertrand M, Duflo E, Mullainathan S.** 2002. "How Much Should We Trust Differences-in-Differences Estimates?." *The Quarterly Journal of Economics* 119(1):249-75.

[12] **Biddle JE, Hamermesh DS.** 1990. "Sleep and the Allocation of Time." *Journal of Political Economy* 98(5, Part 1):922-43.

[13] **Billings ME, Gold DR, Leary PJ, Szpiro A, Aaron CP, Kaufman JD, Redline SS.** 2017. "Relationship of Air Pollution to Sleep Disruption: The Multi-ethnic Study

of Atherosclerosis (MESA) Sleep and MESA-Air Studies." Paper presented at American Thoracic Society Conference.

[14] **Bollen J, Mao H, Zeng X.** 2011. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2(1):1-8.

[15] **Cameron AC, Gelbach JB, Miller DL.** 2008. "Bootstrap-based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics* 90(3):414-27.

[16] **Cameron AC, Miller DL.** 2015. "A Practitioner's Guide to Cluster-robust Inference." *Journal of Human Resources* 50(2):317-72.

[17] **Cappuccio F, Miller MA, Lockley SW (Eds.).** 2010. "Sleep, Health, and Society: From Aetiology to Public Health." *Oxford University Press, USA*.

[18] **Chan M, Wu X, Hao Y, Xi R, Jin T.** 2012. "Microblogging, Online Expression, and Political Efficacy among Young Chinese citizens: The Moderating Role of Information and Entertainment Needs in the Use of Weibo." *Cyberpsychology, Behavior, and Social Networking* 15(7):345-9.

[19] **Chang T, Zivin J, Gross T, Neidell M.** 2016. "Particulate Pollution and the Productivity of Pear Packers." *American Economic Journal: Economic Policy* 8(3):141-69.

[20] **Chen X and Ye J.** 2015. "When the Wind Blows: Spatial Spillover Effects of Urban Air Pollution." Environment for Development Discussion Paper Series, July 2015, EfD DP 15-15.

[21] **Chen Y, Jin GZ, Kumar N, Shi G.** 2012. "Gaming in Air Pollution Data? Lessons from China." *The B.E. Journal of Economic Analysis & Policy* 12(3).

[22] **Chen YL, Yang XJ, Shi CF.** 2004. "Elementary Study on Sleep Quality and Influence Factors for Common Population in Xiamen, China (2004)." *Strait Journal of Preventive Medicine* 6:002.

[23] **Chiu C, Lin D, Silverman A.** 2012. "China's Social-media Boom." *McKinsey & Company*, May 2012. https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/chinas-social-media-boom.

[24] **Choi H, Varian H.** 2012. "Predicting the Present with Google Trends. *Economic Record* 88(s1):2-9.

[25] **Chung KF, Cheung MM.** 2008. "Sleep-wake Patterns and Sleep Disturbance among Hong Kong Chinese Adolescents." *Sleep* 31(2):185-94.

[26] **Currie J, Hanushek EA, Kahn EM, Neidell M, Rivkin SG.** 2009. "Does Pollution Increase School Absences?." *Review of Economics and Statistics* 91(4):682-94.

[27] **Daley M, Morin CM, LeBlanc M, Gregoire JP, Savard J, Baillargeon L.** 2009. "Insomnia and its Relationship to Health-care Utilization, Work absenteeism, Productivity and Accidents." *Sleep Medicine* 10(4):427-38.

[28] **Delfino M, Holt EW, Taylor CR, Wittenberg E, Qureshi AA.** 2008. "Willingness-to-pay Stated Preferences for 8 Health-related Quality-of-life Domains in Psoriasis: A Pilot Study." *Journal of the American Academy of Dermatology* 59(3):439-47.

[29] **DeLuca KM, Brunner E, Sun Y.** 2016. "Weibo, WeChat, and the Transformative Events of Environmental Activism on China's Wild Public Screens." *International Journal of Communication* 10:321-39.

[30] **Diekelmann S, Born J.** 2010. "The Memory Function of Sleep." *Nature Reviews Neuroscience* 11(2):114.

[31] **Driscoll JC, Kraay AC.** 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *The Review of Economics and Statistics* 80(4):549-60.

[32] **Ebenstein A, Lavy V, Roth S.** 2016. "The Long-run Economic Consequences of High-stakes Examinations: Evidence from Transitory Variation in Pollution." *American Economic Journal: Applied Economics* 8(4):36-65.

[33] **EPA (U.S. Environmental Protection Agency).** 1996. "Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information." OAQPS Staff Paper.

[34] **Esarey J, Menger A.** 2015. "Practical and Effective Approaches to Dealing with Clustered Data." Department of Political Science, Rice University.

[35] **Fang L, Clausen G, Fanger PO.** 1998. "Impact of Temperature and Humidity on the Perception of Indoor Air Quality." *Indoor Air* 8(2):80-90.

[36] **Finlayson-Pitts BJ, Pitts JN.** 1997. "Tropospheric Air Pollution: Ozone, Airborne Toxics, Polycyclic Aromatic Hydrocarbons, and Particles." *Science* 276(5315):1045-51.

[37] **Fu S, Viard VB, Zhang P.** 2017. "Air Quality and Manufacturing Firm Productivity: Comprehensive Evidence from China." Unpublished.

[38] **Gao Q, Abel F, Houben GJ, Yu Y.** 2012. "A Comparative Study of Users? Microblogging Behavior on Sina Weibo and Twitter." Presented at International Conference on User Modelling, Adaptation, and Personalization, (pp. 88-101).

[39] **Gayo-Avello D, Metaxas PT, Mustafaraj E.** 2011. "Limits of Electoral Predictions using Twitter." Presented at Proceedings of the Fifth International Association for the Advancement of Artificial Intelligence (AAAI) Conference on Weblogs and Social Media.

[40] **Gerber MS.** 2014. "Predicting Crime Using Twitter and Kernel Density Estimation." *Decision Support Systems* 61:115-25.

[41] **Gendron-Carrier N, Gonzalez-Navarro M, Polloni S, Turner MA.** 2017. "Subways and Urban Air Pollution." Brown University: Providence RI Working Paper.

[42] **Gilliland FD, Berhane K, Rappaport EB, Thomas DC, Avol E, Gauderman WJ, London SJ, Margolis HG, McConnell R, Islam KT, Peters JM.** 2001. "The Effects of Ambient Air Pollution on School Absenteeism due to Respiratory Illnesses." *Epidemiology* 12(1):43-54.

[43] **Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ.** 2010. "Predicting Consumer Behavior with Web Search." *Proceedings of the National Academy of Sciences* 107(41):17486-90.

[44] **Gordon AM, Chen S.** 2014. "The Role of Sleep in Interpersonal Conflict: Do Sleepless Nights Mean Worse Fights?." *Psychological and Personality Science* 5(2):168-75.

[45] **Hamilton NA, Nelson CA, Stevens N, Kitzman H.** 2007. "Sleep and Psychological Well-being." *Social Indicators Research* 82(1):147-63.

[46] **Harden JJ.** 2011. "A Bootstrap Method for Conducting Statistical Inference with Clustered Data." *State Politics & Policy Quarterly* 11(2):223-46.

[47] **Herrnstadt E, Heyes A, Muehlegger E, Saberian S.** 2016. "Air Pollution as a Cause of Violent Crime: Evidence from Los Angeles and Chicago." Unpublished.

[48] **Herrnstadt E, Muehlegger E.** 2014. "Weather, Salience of Climate Change and Congressional Voting." *Journal of Environmental Economics and Management* 68(3):435-48.

[49] **Hoechle D.** 2007. "Robust Standard Errors for Panel Regressions with Cross-sectional Dependence." *Stata Journal* 7(3):281.

[50] **Huber R, Ghilardi MF, Massimini M, Tononi G.** 2004."Local Sleep and Learning." *Nature* 430(6995):78.

[51] **Ireland JL, Culpin V.** 2006. "The Relationship between Sleeping Problems and Aggression, Anger, and Impulsivity in a Population of Juvenile and Young Offenders." *Journal of Adolescent Health* 38(6):649-55.

[52] **Killgore WD, Balkin TJ, Wesensten NJ.** 2006. "Impaired Decision Making Following 49 h of Sleep Deprivation." *Journal of Sleep Research* 15(1):7-13.

[53] **Knittel CR, Miller DL, Sanders NJ.** 2016. "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health." *Review of Economics and Statistics* 98(2):350-66.

[54] **Lavaine E, Neidell M.** 2017. "Energy Production and Health Externalities: Evidence from Oil Refinery Strikes in France." *Journal of the Association of Environmental and Resource Economists* 4(2):447-77.

[55] **Li M, Ren X, YU Y and Zhou L.** 2016. "Spatiotemporal Pattern of Ground-level Fine Particulate Matter ($PM_{2.5}$) Pollution in Mainland China." *China Environmental Science* 36(3):641-50.

[56] **Luo J, Zhu G, Zhao Q, Guo Q, Meng H, Hong Z, Ding D.** 2013. "Prevalence and Risk Factors of Poor Sleep Quality among Chinese Elderly in an Urban Community: Results from the Shanghai Aging Study." *PloS One* 8(11):e81261.

[57] **Maquet P.** 2001. "The Role of Sleep in Learning and Memory." *Science* 294(5544): 1048-52.

[58] **Meijer AM, Reitz E, Deković M, Van Den Wittenboer GL, Stoel RD.** 2010. "Longitudinal Relations between Sleep Quality, Time in Bed and Adolescent Problem Behavior." *Journal of Child Psychology and Psychiatry* 51(11):1278-86.

[59] **Mendell MJ, Heath GA.** 2005. "Do Indoor Pollutants and Thermal Conditions in Schools Influence Student Performance? A Critical Review of the Literature." *Indoor Air* 15(1):27-52.

[60] **Moretti E, Neidell M.** 2011. "Pollution, Health, and Avoidance Behavior Evidence from the Ports of Los Angeles." *Journal of Human Resources* 46(1):154-75.

[61] **National Research Council.** 1993. "Protecting Visibility in National Parks and Wilderness Areas." *National Academies Press.*

[62] **Norman J.** 1988. "Chinese." *Cambridge University Press.*

[63] **Okamoto-Mizuno K, Mizuno K.** 2012. "Effects of Thermal Environment on Sleep and Circadian Rhythm." *Atmospheric Environment* 31(1):14.

[64] **Pollinger M.** 2014. "Individuals Willingness to Pay for Health and Wellness in the Built Environment." *Journal of Environmental and Resource Economics at Colby* 1(01): 52.

[65] **Ransom MR, Pope CA.** 1992. "Elementary school absences and PM_{10} pollution in Utah Valley." *Environmental Research* 58(1):204-19.

[66] **Rechtschaffen A.** 1971. "The Control of Sleep." *Human Behavior and Control* p.88.

[67] **Rosekind MR, Gregory KB, Mallis MM, Brandt SL, Seal B, Lerner D.** 2010. "The Cost of Poor Sleep: Workplace Productivity Loss and Associated Costs." *Journal of Occupational and Environmental Medicine* 52(1):91-8.

[68] **Sager L.** 2016. "Estimating the Effect of Air Pollution on Road Safety Using Atmospheric Temperature Inversions." Grantham Research Institute on Climate Change and the Environment Working Paper, 251.

[69] **Schlenker W, Walker WR.** 2016. "Airports, Air pollution, and Contemporaneous Health." *The Review of Economic Studies* 83(2):768-809.

[70] **Shao M, Tang X, Zhang Y, Li W.** 2006. "City Clusters in China: Air and Surface Water Pollution." *Frontiers in Ecology and the Environment* 4(7):353-61.

[71] **Steptoe A, O'Donnell K, Marmot M, Wardle J.** 2008 "Positive Affect, Psychological Well-being, and Good Sleep." *Journal of Psychosomatic Research* 64(4):409-15.

[72] **Stoerk T.** 2016. "Statistical Corruption in Beijing's Air Quality Data Has Likely Ended in 2012." *Atmospheric Environment* 127:365-71.

[73] **Strøm-Tejsen P, Zukowska D, Wargocki P, Wyon DP.** 2016. "The Effects of Bedroom Air Quality on Sleep and Next-day Performance." *Indoor Air* 26(5):679-86.

[74] **Sullivan J.** 2012. "A Tale of Two Microblogs in China." *Media, Culture & Society* 34 (6):773-83.

[75] **Sun W, Yu Y, Yuan J, Li C, Liu T, Lin D, Lau A, Zhong C, Xu T, Shan G.** 2015. "Sleep Duration and Quality among Different Occupations — China National Study." *PloS One* 10(3):e0117700.

[76] **Valent F, Bartolomeo SD, Marchetti R, Sbrojavacca R, Barbone F.** 2010. "A Case-crossover Study of Sleep and Work Hours and the Risk of Road Traffic Accidents." *Sleep* 33(3):349-54.

[77] **Van Someren EJ.** 2006. "Mechanisms and Functions of Coupling between Sleep and Temperature Rhythms." *Progress in Brain Research* 153:309-24.

[78] **Wang J, Ogawa S.** 2015. "Effects of Meteorological Conditions on $PM_{2.5}$ Concentrations in Nagasaki, Japan." *International Journal of Environmental Research and Public Health* 12(8):9089-101.

[79] **Wolkoff P.** 1998. "Impact of Air Velocity, Temperature, Humidity, and Air on Long-term VOC Emissions from Building Products." *Atmospheric Environment* 32(14):2659-68.

[80] **Wooldridge JM.** 2010. "Econometric Analysis of Cross Section and Panel Data." *MIT Press.*

[81] **Yu Y, Lu BS, Wang B, Wang H, Yang J, Li Z, Wang L, Liu X, Tang G, Xing H, Xu X.** 2007. "Short Sleep Duration and Adiposity in Chinese Adolescents." *Sleep* 30 (12):1688-97.

[82] **Zahran S, Iverson T, McElmurry SP, Weiler S.** 2017."The Effect of Leaded Aviation Gasoline on Blood Lead in Children." *Journal of the Association of Environmental and Resource Economists* 4(2):575-610.

[83] **Zammit GK, Joish VN, Kong MC, Balkrishnan R, Lerner D, Rosekind M.** 2010. "Impact of Nighttime Awakenings on Worker Productivity and Performance." *Journal of Occupational and Environmental Medicine* 52(2):513-8.

[84] **Zanobetti A, Redline S, Schwartz J, Rosen D, Patel S, O'Connor GT, Lebowitz M, Coull BA, Gold DR.** 2010. "Associations of PM_{10} with Sleep and Sleep-disordered Breathing in Adults from Seven US Urban Areas." *American Journal of Respiratory and Critical Care Medicine* 182(6):819-25.

[85] **Zhang J, Lioy PJ.** 1994. "Ozone in Residential Air: Concentrations, I/O Ratios, Indoor Chemistry, and Exposures." *Indoor Air* 4(2):95-105.

[86] **Zheng S, Cao J, Kahn ME, Sun C.** 2014. "Real Estate Valuation and Cross-boundary Air Pollution Externalities: Evidence from Chinese cities." *The Journal of Real Estate Finance and Economics* 48(3):398-414.

[87] **Zivin JG, Neidell M.** 2012. "The Impact of Pollution on Worker Productivity." *The American Economic Review* 102(7):3652-73.

Table 1: Summary Statistics

	Mean	Std. Dev.	Min	Max
Sleeplessness Index	276.660	168.088	54	1258
AQI Index	96.465	55.557	14	475
$PM_{2.5}(\mu g/m^3)$	67.328	47.843	5.2	543.5
$PM_{10}(\mu g/m^3)$	100.981	62.371	0	537.6
$CO(mg/m^3)$	1.143	0.583	0.222	8.141
$NO_2(\mu g/m^3)$	46.999	19.664	4.9	175.8
$SO_2(\mu g/m^3)$	22.944	19.521	2	212.6
$O_3(\mu g/m^3)$	106.309	61.566	3	357
Avg Temperature($^\circ C$)	17.210	9.032	-6.7	35.2
Max Temperature($^\circ C$)	21.520	9.409	-3.9	41.1
Min Temperature($^\circ C$)	13.514	9.145	-11.2	30.6
Avg Humidity(%)	69.300	17.321	0	99
Max Humidity(%)	85.535	13.193	10	100
Min Humidity(%)	52.317	21.192	0	99
Sea-level Pressure(Hpa)	1016.095	9.277	990.5	1040.2
Wind Speed(Km/h)	7.866	3.490	0.9	27.9
Cloud Cover(-/8)	5.367	2.646	0	8
Precipitation(mm)	3.743	11.250	0	173.1

Notes: The dataset contains daily data from ten target cities for 2014 to 2015. The total number of the observations is 7300.

Table 2: Air Quality Index (AQI) and Pollutant Concentrations

	Level	Description	AQI	$PM_{2.5}$ (24hr) ($\mu g/m^3$)	CO (24hr) (mg/m^3)	NO_2 (24hr) ($\mu g/m^3$)	Number of Days (AQI)	Percent
Low	I	Excellent	0-50	0-35	0-2	0-40	1129	15.47%
	II	Good	51-100	36-75	2.1-4	41-80	3593	49.22%
Medium	III	Light Polluted	101-150	76-115	4.1-14	81-180	1555	21.30%
		Moderately Polluted	151-200	115-150	14.1-24	181-280	615	8.42%
High	IV	Heavily Polluted	201-300	151-250	24.1-36	281-565	335	4.59%
	V	Severely Polluted	301-500	251-500	36.1-60	566-940	73	1.00%

Notes: The table maps each pollutant ambient concentration corresponding to AQI categories. Classification principles are taken from *Technical Regulation on Ambient Air Quality Index HJ 633-2012*. Level I and II do not have health implications, suitable for outdoor activities. Higher level of pollutant leads to higher risk of breathing or heart problems. Outdoor exercise should be reduced. Level V may induce respiratory diseases, and outdoor exposure is avoid for elderly and sick people. The last two columns summarize the number of days under each level (measured by daily mean) and individual percentage among 7300 observations (ten cities within two years). The last two columns report the number of days and corresponding percentage of days falling into each category in sample.

Table 3: Correlation between Pollutants

	AQI	$PM_{2.5}$	PM_{10}	CO	NO_2	SO_2	O_3
AQI	1.000						
$PM_{2.5}$	0.977	1.000					
PM_{10}	0.898	0.850	1.000				
CO	0.734	0.752	0.683	1.000			
NO_2	0.651	0.652	0.680	0.624	1.000		
SO_2	0.565	0.554	0.618	0.663	0.598	1.000	
O_3	-0.028	-0.073	0.001	-0.266	-0.089	-0.180	1.000

Notes: The table displays covariance matrix of pollutants in the dataset.

Table 4: Air Quality and Sleeplessness — OLS

Independent Variable	Dependent Variable Ln(Sleepless)				
Daily Pollutant	City FEs	Temporal Controls			Weather Covariates
	(1)	(2)	(3)	(4)	(5)
Panel A: AQI	0.097***	0.037**	0.035**	0.035**	0.043***
	(0.030)	(0.017)	(0.017)	(0.017)	(0.014)
Panel B: PM2.5	0.105***	0.047**	0.044**	0.045**	0.049***
	(0.037)	(0.022)	(0.021)	(0.022)	(0.017)
Observations	7300	7300	7300	7300	6839
Additional Control					
City FEs	Y	Y	Y	Y	Y
Year_season FEs	N	Y	Y	Y	Y
Day of Week	N	N	Y	Y	Y
Holiday	N	N	N	Y	Y
Weather Covariates	N	N	N	N	Y

Notes: Dependent variable is log form of Sleeplessness Index. Data collection period runs from 11pm to 7am. Independent variable of interest is daily average measure of specific pollutant. All estimators have been adjusted into percentage by multiplying 100. Temporal controls include day of week and holiday fixed effects, as well as year_season fixed effects. Weather controls contain temperature, humidity, precipitation, wind speed and sea-level pressure. Temperature and humidity are measured by the way of bins (5 degree C indicators for average temperature, 20 percent indicators for average humidity); continuous maximum and minimum values are also included in the estimation. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table 5: Air Quality and Sleeplessness — IV

2SLS	City FEs		Temporal Controls		Weather Covariates	
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: AQI			
First Stage[(a)]						
Instrumental AQI t	0.803***	0.298***	0.640***	0.247***	0.624***	0.277***
	(0.055)	(0.038)	(0.055)	(0.036)	(0.064)	(0.035)
Instrumental AQI lagged t-1		0.583***		0.532***		0.489***
		(0.063)		(0.059)		(0.062)
F-statistics	217.03	302.59	136.61	275.14	95.29	184.7
Second Stage[(b)]						
Instrumented AQI	0.378***	0.371***	0.249***	0.234***	0.198**	0.196**
	(0.072)	(0.068)	(0.087)	(0.078)	(0.089)	(0.084)
			Panel B: PM2.5			
First Stage[(a)]						
Instrumental $PM_{2.5}$ t	0.806***	0.316***	0.630***	0.257***	0.586***	0.258***
	(0.050)	(0.052)	(0.047)	(0.061)	(0.069)	(0.045)
Instrumental $PM_{2.5}$ lagged t-1		0.569***		0.509***		0.463***
		(0.083)		(0.070)		(0.068)
F-statistics	260.02	325.42	177.71	221.61	72.35	108.94
Second Stage[(b)]						
Instrumented	0.385***	0.380***	0.279***	0.262***	0.228**	0.224**
	(0.089)	(0.085)	(0.105)	(0.097)	(0.115)	(0.110)
Observations	7023	6930	7023	6930	6567	6475
Additional Control						
City FEs	Y	Y	Y	Y	Y	Y
Year_season FEs	N	N	Y	Y	Y	Y
Day of Week	N	N	Y	Y	Y	Y
Holiday	N	N	Y	Y	Y	Y
Weather Covariates	N	N	N	N	Y	Y

Notes: (a) Dependent variable in the first stage is daily-mean pollutant of target city, and independent variable is daily weighted average pollution of surrounding cities ($100km < d_{ij} < 200km$) from upwind direction (within 90 degree to the wind). (b) Second stage reports the results regressing log form of Sleeplessness Index on the instrumented daily pollution with estimators being adjusted into percentage by multiplying 100. Column (2), (4) and (6) incorporate day before as additional instrument. Temporal controls include day of week, holiday fixed effects and year_season fixed effects. Weather controls contain temperature, humidity, precipitation, wind speed and sea-level pressure. Temperature and humidity are measured by the way of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table 6: Robustness — IV with 60 Degree Wind Angle Inclusion

2SLS	AQI		PM2.5	
	(1)	(2)	(3)	(4)
First Stage[(a)]				
Instrumental Pollutant t	0.532***	0.249***	0.506***	0.234***
	(0.087)	(0.043)	(0.089)	(0.058)
Instrumental Pollutant lagged t-1		0.434***		0.414***
		(0.070)		(0.081)
F-statistics	37.5	77.24	32.56	62.93
Second Stage[(b)]				
Instrumented Pollutant	0.261**	0.221***	0.252**	0.258**
	(0.097)	(0.089)	(0.122)	(0.115)
Observations	6567	6475	6567	6475
Additional Control				
City FEs	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y

Notes: (a) Dependent variable in the first stage is daily-mean pollutant of target city, and independent variable is daily weighted average pollution of surrounding cities ($100km < d_{ij} < 200km$) from upwind direction (within 60 degree to the wind). (b) Second stage reports the results regressing log form of Sleeplessness Index on the instrumented daily pollution. Column (2) and (4) incorporate day before as additional instrument. Temporal controls include day of week and holiday fixed effects. Weather controls contain temperature, humidity, precipitation, wind speed and sea-level pressure. Temporal controls include day of week, holiday fixed effects and year_season fixed effects. Weather controls contain temperature, humidity, precipitation, wind speed and sea-level pressure. Temperature and humidity are measured by the way of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table 7: Reduced Form and Drop Weighting Scheme

	OLS	IV		Reduced From	
	Local Pollutant	Pollutant (Weighted Average)	Pollutant (Arithmetic Average)	Pollutant (Weighted Average)	Pollutant (Arithmetic Average)
	(1)	(2)	(3)	(4)	(5)
Panel A: AQI	0.043***	0.198**	0.185**	0.124**	0.128**
	(0.014)	(0.089)	(0.089)	(0.056)	(0.062)
Panel B: PM2.5	0.049***	0.228**	0.212*	0.134**	0.135*
	(0.017)	(0.115)	(0.117)	(0.067)	(0.075)
Observations	6839	6567	6567	6567	6567
Additional Controls					
City FEs	Y	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y	Y

Notes: Column (1) repeats the OLS results of Column (5) in Table 4. Column (2) repeats the second stage results under Column (5) in Table 5. Column (3) re-constructs the instrument with arithmetic average among upwind cities. Column (4) presents the results of reduced form regressing the log form of daily Sleeplessness Index on daily weighted average pollutant of peripheral cities ($100km < d_{ij} < 200km$) from upwind direction (within 90 degree to the wind). Column (5) tests arithmetic average among upwind cities. All the regressions include city fixed effects, temporal controls (day of week, holiday fix effects and year_season fixed effects) and weather controls (average temperature bins, max and min temperature, precipitation, sea-level pressure, wind speed, average humidity bins, max and min humidity). Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table 8: Precipitation Exclusion — OLS

	Full	Clear Nights	Zero Rain Days
	(1)	(2)	(3)
Panel A: AQI	0.043***	0.043***	0.045**
	(0.014)	(0.017)	(0.019)
Panel B: PM2.5	0.049***	0.050**	0.053**
	(0.017)	(0.021)	(0.023)
Observations	6839	5620	4179
Additional Controls			
City FEs	Y	Y	Y
Temporal Controls	Y	Y	Y
Weather Covariates	Y	Y	Y

Notes: Dependent variable is log form of Sleeplessness Index. Independent variable is city daily-mean value of specific pollutant. Column (1) displays the results for all observations replicating the results under Column (5) in Table 4. Column (2) excludes days with precipitation from 11pm to 7am. Column (3) excludes days with precipitation from 12pm to 7am on the following day. All the regressions include city fixed effects, temporal controls and weather covariates. Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table 9: Precipitation Exclusion — IV

	Full	Clear Nights	Zero Rain Days
	(1)	(2)	(3)
		Panel A: AQI	
First Stage			
Instrumental AQI t	0.277***	0.252***	0.218***
	(0.035)	(0.042)	(0.045)
Instrumental AQI lagged t-1	0.489***	0.490***	0.489***
	(0.062)	(0.070)	(0.062)
F-Statistic	184.7	145.45	64.66
Second Stage			
Instrumented Pollutant	0.196**	0.173**	0.186**
	(0.084)	(0.085)	(0.092)
		Panel B: PM2.5	
First Stage[(a)]			
Instrumental $PM_{2.5}$ t	0.258***	0.267***	0.227***
	(0.045)	(0.043)	(0.046)
Instrumental $PM_{2.5}$ lagged t-1	0.463***	0.430***	0.422***
	(0.068)	(0.073)	(0.072)
F-Statistic	108.94	100.65	38.14
Second Stage[(b)]			
Instrumented Pollutant	0.224**	0.202*	0.227*
	(0.110)	(0.112)	(0.123)
Observations	6475	5356	3986
Additional Controls			
City FEs	Y	Y	Y
Temporal Controls	Y	Y	Y
Weather Covariates	Y	Y	Y

Notes: Column (1) displays IV results for all observations, reprint the regressions under Column (6) in Table 5. Column (2) excludes the days with snowy or rainy nights. Column (3) limits to clear days without rain or snow in the daytime or nighttime. (a) First stage reports the results regressing city daily-mean pollutant on imported pollutants from source cities. (b) Dependent variable in the second stage is log form of Sleeplessness Index, and independent variable is daily average measure of specific pollutant. All the regressions include city fixed effects, temporal controls and weather covariates. Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table 10: Alternative Standard Errors — OLS

	Pairs Cluster Bootstrap (1)	Wild Cluster Bootstrap (2)	Driscoll-Kraay Spatial Correlation (3)
Panel A: AQI	0.043***	0.043**	0.043**
	[0.002]	[0.03]	[0.02]
Panel B: PM2.5	0.049***	0.049*	0.049**
	[0.004]	[0.06]	[0.022]
Observations	6839	6839	6839
Additional Controls			
City FEs	Y	Y	Y
Temporal Controls	Y	Y	Y
Weather Covariates	Y	Y	Y

Notes: Dependent variable is log form of Sleeplessness Index. Data collection period runs from 11pm to 7am. Independent variable of interest is daily average measure of specific pollutant. Each reported coefficient is derived from a separate regression. All regressions include city fixed effects, temporal controls (day of week fixed effect and holiday fixed effect) and weather controls (average temperature bins, max and min temperature, precipitation, sea-level pressure, wind speed, average humidity bins, max and min humidity). Column (1) repeats the results of Column (5) in Table 4. Column (2) implements the wild bootstrap procedure as described in Cameron et al. (2008), an alternative way to to address the small number of clusters. Column (3) follows Driscoll and Kraay (1998) to adjust for spatial correlation. P-values are shown in brackets. (* significant at 10%, ** significant at 5%, *** significant at 1%)

Table 11: Alternative Standard Errors — IV

	Pairs Cluster Bootstrap (1)	Wild Cluster Bootstrap (2)	Driscoll-Kraay Spatial Correlation (3)
		Panel A: AQI	
First Stage			
Instrumental AQI t	0.277***	0.277***	0.277***
	[0.00]	[0.00]	[0.00]
Instrumental AQI lagged t-1	0.489***	0.489***	0.489***
	[0.00]	[0.00]	[0.00]
F-Statistic	184.7	-	78.06
Second Stage			
Instrumented Pollutant	0.196**	0.196**	0.196***
	[0.019]	[0.04]	[0.014]
		Panel B: PM2.5	
First Stage			
Instrumental $PM_{2.5}$ t	0.258***	0.258***	0.258***
	[0.00]	[0.00]	[0.00]
Instrumental $PM_{2.5}$ lagged t-1	0.463***	0.463***	0.463***
	[0.00]	[0.00]	[0.00]
F-Statistic	108.94	-	71.24
Second Stage			
Instrumented Pollutant	0.224**	0.224*	0.224**
	[0.041]	[0.10]	[0.021]
Observations	6475	6475	6475
Additional Controls			
City FEs	Y	Y	Y
Temporal Controls	Y	Y	Y
Weather Covariates	Y	Y	Y

Notes: Dependent variable in the first stage is daily-mean pollutant of local city, and independent variable is daily weighted average pollution of source cities. Second stage reports the results regressing log Sleeplessness Index on the instrumented daily pollution. Column (1) repeats the IV results of Column (6) in Table 5. Pairs bootstrapped standard errors clustered at city level are reported in parentheses. Column (2) implements the wild bootstrap procedure as described in Cameron et al. (2008), an alternative way to to address the small number of clusters. Column (3) follows Driscoll and Kraay (1998) to consider spatial correlation. P-values are shown in brackets. (* significant at 10%, ** significant at 5%, *** significant at 1%)

Table 12: Joint Estimation

	OLS		2SLS		
	Single Estimation	Joint Estimation	Single Estimation	Joint Estimation (Schlenker and Walker 2016)	Joint Estimation (Moretti and Neidell 2011)
	(1)	(2)	(3)	(4)	(5)
$PM_{2.5}$	0.049***	0.038	0.224**	0.287*	0.434*
	(0.017)	(0.025)	(0.110)	(0.156)	(0.258)
CO	4.138**	2.269	15.292	5.961	29.949
	(1.851)	(1.545)	(11.160)	(11.486)	(27.885)
NO_2	0.076	-0.028	0.086	-0.416	-0.130
	(0.055)	(0.082)	(0.255)	(0.361)	(0.535)
O_3	0.002	-0.002	-	-	-
	(0.019)	(0.020)	-	-	-
Additional Controls					
City FEs	Y	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y	Y
Year_season FEs	Y	Y	Y	Y	Y

Notes: Column (1) and Column (3) repeat the OLS and IV estimations from the preferred specification in Table 4 and 5. Joint estimations that include $PM_{2.5}$, CO, NO_2 and O_3 together are reported in Column (2), Column (4) and Column (5). Column (4) regresses Sleeplessness Index on different instrumented pollutants together. Column (5) controls co-pollution when making instrument and report each second stage estimate one by one. The instruments are as used in the even columns in Table 5. All the regressions include city fixed effects, temporal controls and weather controls. Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Online Appendix — Not for Publication

Table A1: Summary Statistics by City

	Beijing	Changsha	Chongqing	Guangzhou	Hangzhou
Sleeplessness Index	591.530	313.432	160.747	385.870	234.129
	(224.858)	(65.926)	(66.828)	(151.775)	(89.351)
AQI Index	117.011	92.921	84.159	64.862	83.997
	(77.852)	(51.611)	(45.162)	(28.295)	(37.603)
$PM_{2.5}(\mu g/m^3)$	82.306	67.490	58.904	43.214	58.422
	(70.573)	(43.359)	(37.237)	(23.068)	(31.218)
$PM_{10}(\mu g/m^3)$	107.897	82.563	89.960	64.545	89.207
	(76.330)	(44.099)	(48.667)	(30.065)	(43.582)
$CO(mg/m^3)$	1.282	1.025	1.129	0.981	0.896
	(1.007)	(0.342)	(0.269)	(0.259)	(0.251)
$NO_2(\mu g/m^3)$	51.799	38.506	40.302	45.739	46.556
	(24.33)	(16.257)	(11.824)	(17.635)	(16.005)
$SO_2(\mu g/m^3)$	16.683	20.870	19.777	14.573	17.748
	(19.543)	(10.55)	(11.739)	(6.36)	(9.13)
$O_3(\mu g/m^3)$	111.401	87.281	75.725	111.007	109.007
	(72.81)	(46.496)	(55.83)	(60.831)	(56.888)
Avg Temperature($^\circ C$)	13.977	18.134	19.201	22.140	17.576
	(10.768)	(8.222)	(7.318)	(6.161)	(8.261)
Max Temperature($^\circ C$)	19.077	21.778	22.626	26.807	21.691
	(11.188)	(9.01)	(8.393)	(6.462)	(8.804)
Min Temperature($^\circ C$)	9.111	15.278	16.612	18.866	14.323
	(10.556)	(7.92)	(6.634)	(6.35)	(8.225)
Avg Humidity(%)	52.444	74.923	76.393	76.284	73.362
	(20.008)	(15.668)	(11.223)	(10.191)	(14.349)
Max Humidity(%)	74.330	88.351	89.288	89.505	87.982
	(19.492)	(10.041)	(7.447)	(6.637)	(9.183)
Avg Humidity(%)	34.534	59.971	62.764	58.426	57.036
	(19.562)	(21.477)	(15.954)	(16.215)	(19.707)
Sea-level Pressure(Hpa)	1016.919	1015.715	1013.733	1013.420	1016.596
	(10.045)	(9.074)	(9.128)	(6.994)	(9.086)
Wind Speed(Km/h)	7.815	8.160	4.992	8.318	7.812
	(2.928)	(4.012)	(1.304)	(3.265)	(2.652)
Cloud Cover(-/8)	5.179	5.966	6.227	5.951	5.455
	(2.588)	(2.496)	(2.347)	(2.199)	(2.674)
Precipitation(mm)	1.375	4.007	4.485	7.131	5.136
	(5.941)	(10.063)	(10.788)	(17.58)	(11.491)

	Nanjing	Shanghai	Tianjin	Wuhan	Zhengzhou
Sleeplessness Index	219.438	293.386	154.270	227.869	185.927
	(85.274)	(104.944)	(57.28)	(89.196)	(77.087)
AQI Index	95.607	78.511	112.393	105.893	129.292
	(46.566)	(39.397)	(62.051)	(54.571)	(62.349)
$PM_{2.5}(\mu g/m^3)$	65.192	52.990	78.145	75.385	91.230
	(39.568)	(33.372)	(54.121)	(50.426)	(55.595)
$PM_{10}(\mu g/m^3)$	110.119	72.271	127.135	107.816	158.293
	(58.921)	(40.535)	(72.69)	(56.396)	(74.298)
$CO(mg/m^3)$	0.938	0.838	1.518	1.156	1.668
	(0.344)	(0.286)	(0.794)	(0.394)	(0.642)
$NO_2(\mu g/m^3)$	50.383	44.737	47.963	51.146	52.861
	(19.107)	(19.976)	(23.855)	(20.599)	(18.435)
$SO_2(\mu g/m^3)$	21.058	17.301	38.281	25.998	37.147
	(12.024)	(9.941)	(35.229)	(14.982)	(26.806)
$O_3(\mu g/m^3)$	113.629	116.011	91.389	156.686	90.956
	(59.537)	(48.703)	(53.572)	(67.612)	(50.718)
Avg Temperature($^\circ C$)	16.535	17.040	14.247	17.019	16.228
	(8.518)	(8.155)	(10.92)	(8.652)	(9.505)
Max Temperature($^\circ C$)	20.599	20.566	19.284	21.815	20.957
	(8.729)	(8.345)	(11.213)	(8.831)	(9.915)
Min Temperature($^\circ C$)	13.057	13.955	9.380	12.940	11.617
	(8.636)	(8.436)	(10.667)	(9.026)	(9.341)
Avg Humidity(%)	72.692	73.001	56.551	78.345	59.008
	(14.391)	(12.869)	(17.46)	(10.536)	(18.051)
Max Humidity(%)	88.232	86.761	78.307	94.473	78.119
	(9.324)	(8.948)	(15.683)	(5.923)	(16.255)
Min Humidity(%)	55.800	57.933	37.600	57.381	41.736
	(19.908)	(18.546)	(19.073)	(18.162)	(19.241)
Sea-level Pressure(Hpa)	1017.052	1017.075	1017.158	1016.110	1017.178
	(9.25)	(8.966)	(9.992)	(9.235)	(9.727)
Wind Speed(Km/h)	9.422	9.295	9.986	5.841	7.021
	(3.784)	(3.293)	(3.698)	(3.088)	(2.673)
Cloud Cover(-/8)	5.267	4.978	3.778	5.993	4.874
	(2.645)	(2.717)	(2.776)	(2.342)	(2.707)
Precipitation(mm)	4.241	4.119	1.449	3.789	1.827
	(14.291)	(11.656)	(6.347)	(11.261)	(6.739)

Notes: The table lists the sample means at daily level. Standard deviations are shown in parentheses.

Table A2: Average Correlation among Monitoring Stations within Each City

City	Average Correlation AQI	Average Correlation $PM_{2.5}$	Average Correlation CO	Average Correlation NO_2	Average Correlation O_3
Beijing	0.885	0.952	0.933	0.881	0.944
Changsha	0.925	0.966	0.544	0.692	0.812
Chongqing	0.831	0.895	0.553	0.672	0.840
Guangzhou	0.845	0.932	0.454	0.795	0.860
Hangzhou	0.803	0.866	0.665	0.719	0.852
Nanjing	0.942	0.972	0.720	0.772	0.880
Shanghai	0.932	0.966	0.773	0.908	0.873
Tianjin	0.881	0.940	0.789	0.908	0.812
Wuhan	0.897	0.947	0.708	0.802	0.724
Zhengzhou	0.905	0.955	0.746	0.834	0.856
Overall Average	0.885	0.939	0.688	0.798	0.845

Notes: The table reports the average pairwise correlations for daily average pollutant levels from all the monitoring stations in each city. The mean values under Beijing are the same as those average values from Table A3 to A7.

Table A3: Pairwise Correlations among Monitoring Stations in Beijing — AQI

Correlation	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9	Station 10	Station 11	Station 12
Station 1	-											
Station 2	0.792	-										
Station 3	0.957	0.816	-									
Station 4	0.976	0.799	0.962	-								
Station 5	0.947	0.811	0.968	0.960	-							
Station 6	0.962	0.826	0.963	0.961	0.953	-						
Station 7	0.928	0.854	0.936	0.927	0.931	0.961	-					
Station 8	0.883	0.818	0.888	0.888	0.897	0.889	0.879	-				
Station 9	0.840	0.857	0.841	0.841	0.844	0.858	0.864	0.915	-			
Station 10	0.776	0.907	0.795	0.784	0.794	0.804	0.840	0.790	0.821	-		
Station 11	0.931	0.841	0.956	0.938	0.959	0.959	0.950	0.899	0.863	0.812	-	
Station 12	0.912	0.839	0.904	0.900	0.892	0.932	0.940	0.865	0.862	0.817	0.910	-
Average	0.885											
Longitude	116.366	116.170	116.434	116.434	116.473	116.361	116.315	116.720	116.644	116.230	116.407	116.225
Latitude	39.867	40.287	39.952	39.875	39.972	39.943	39.993	40.144	40.394	40.195	40.003	39.928

Notes: The table reports the correlations among daily average AQI generated from each monitoring station in Beijing.

Table A4: Pairwise Correlations among Monitoring Stations in Beijing — $PM_{2.5}$

Correlation	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9	Station 10	Station 11	Station 12
Station 1	-											
Station 2	0.868	-										
Station 3	0.968	0.908	-									
Station 4	0.984	0.894	0.992	-								
Station 5	0.963	0.906	0.992	0.989	-							
Station 6	0.971	0.913	0.992	0.990	0.986	-						
Station 7	0.953	0.937	0.978	0.974	0.975	0.988	-					
Station 8	0.948	0.912	0.957	0.960	0.964	0.962	0.960	-				
Station 9	0.918	0.937	0.926	0.926	0.930	0.940	0.949	0.973	-			
Station 10	0.887	0.978	0.921	0.912	0.924	0.923	0.952	0.921	0.934	-		
Station 11	0.957	0.914	0.989	0.985	0.988	0.994	0.986	0.965	0.939	0.925	-	
Station 12	0.958	0.925	0.967	0.970	0.965	0.979	0.983	0.961	0.951	0.933	0.975	-
Average	0.952											
Longitude	116.366	116.170	116.434	116.434	116.473	116.361	116.315	116.720	116.644	116.230	116.407	116.225
Latitude	39.867	40.287	39.952	39.875	39.972	39.943	39.993	40.144	40.394	40.195	40.003	39.928

Notes: The table reports the correlations among daily average $PM_{2.5}$ generated from each monitoring station in Beijing.

Table A5: Pairwise Correlations among Monitoring Stations in Beijing — CO

Correlation	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9	Station 10	Station 11	Station 12
Station 1	-											
Station 2	0.858	-										
Station 3	0.953	0.881	-									
Station 4	0.966	0.872	0.987	-								
Station 5	0.950	0.876	0.980	0.978	-							
Station 6	0.958	0.884	0.989	0.988	0.980	-						
Station 7	0.914	0.885	0.962	0.949	0.946	0.966	-					
Station 8	0.928	0.888	0.932	0.933	0.933	0.927	0.902	-				
Station 9	0.909	0.916	0.913	0.914	0.911	0.914	0.893	0.952	-			
Station 10	0.903	0.944	0.922	0.915	0.916	0.923	0.937	0.911	0.927	-		
Station 11	0.954	0.902	0.977	0.974	0.979	0.981	0.943	0.941	0.922	0.929	-	
Station 12	0.937	0.895	0.942	0.943	0.944	0.951	0.952	0.907	0.911	0.936	0.948	-
Average	0.933											
Longitude	116.366	116.170	116.434	116.434	116.473	116.361	116.315	116.720	116.644	116.230	116.407	116.225
Latitude	39.867	40.287	39.952	39.875	39.972	39.943	39.993	40.144	40.394	40.195	40.003	39.928

Notes: The table reports the correlations among daily average CO generated from each monitoring station in Beijing.

Table A6: Pairwise Correlations among Monitoring Stations in Beijing — NO_2

Correlation	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9	Station 10	Station 11	Station 12
Station 1	-											
Station 2	0.799	-										
Station 3	0.962	0.780	-									
Station 4	0.962	0.794	0.965	-								
Station 5	0.946	0.808	0.967	0.948	-							
Station 6	0.968	0.802	0.961	0.955	0.949	-						
Station 7	0.919	0.750	0.882	0.858	0.886	0.904	-					
Station 8	0.897	0.796	0.894	0.872	0.914	0.881	0.868	-				
Station 9	0.827	0.885	0.780	0.792	0.831	0.817	0.818	0.877	-			
Station 10	0.866	0.941	0.856	0.864	0.884	0.872	0.831	0.876	0.892	-		
Station 11	0.932	0.789	0.970	0.948	0.948	0.932	0.847	0.883	0.773	0.844	-	
Station 12	0.943	0.843	0.933	0.929	0.919	0.935	0.887	0.875	0.826	0.881	0.927	-
Average	0.881											
Longitude	116.366	116.170	116.434	116.434	116.473	116.361	116.315	116.720	116.644	116.230	116.407	116.225
Latitude	39.867	40.287	39.952	39.875	39.972	39.943	39.993	40.144	40.394	40.195	40.003	39.928

Notes: The table reports the correlations among daily average NO_2 generated from each monitoring station in Beijing.

Table A7: Pairwise Correlations among Monitoring Stations in Beijing — O_3

Correlation	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9	Station 10	Station 11	Station 12
Station 1	-											
Station 2	0.900	-										
Station 3	0.974	0.907	-									
Station 4	0.974	0.900	0.983	-								
Station 5	0.968	0.905	0.990	0.976	-							
Station 6	0.969	0.901	0.984	0.972	0.978	-						
Station 7	0.946	0.899	0.972	0.951	0.970	0.969	-					
Station 8	0.937	0.912	0.960	0.950	0.961	0.942	0.936	-				
Station 9	0.884	0.915	0.906	0.900	0.905	0.886	0.885	0.957	-			
Station 10	0.933	0.943	0.948	0.941	0.947	0.942	0.947	0.957	0.942	-		
Station 11	0.953	0.902	0.975	0.963	0.974	0.968	0.970	0.948	0.902	0.949	-	
Station 12	0.969	0.896	0.974	0.965	0.969	0.972	0.970	0.936	0.883	0.947	0.971	-
Average	0.944											
Longitude	116.366	116.170	116.434	116.434	116.473	116.361	116.315	116.720	116.644	116.230	116.407	116.225
Latitude	39.867	40.287	39.952	39.875	39.972	39.943	39.993	40.144	40.394	40.195	40.003	39.928

Notes: The table reports the correlations among daily average O_3 generated from each monitoring station in Beijing.

Table A8: Target City and Source Instrumental Cities

Target Cities	Instrumental Cities	Coordinate (°)	Distance (km)	Location (°)
Beijing	Chengde	40.97N 117.94E	175.93	56.22
39.96N 116.43E	Tangshan	39.62N 118.18E	154.4	100.99
	Tianjin	39.09N 117.20E	113.34	139.13
	Langfang	39.54N 116.68E	48.48	149.24
	Baoding	38.89N 115.47E	140.06	221.9
	Zhangjiakou	40.76N 114.88E	161.34	297.3
Changsha	Yueyang	29.36N 113.31E	126.24	9.64
28.23N 112.94E	Xinyu	27.81N 114.92E	200	101.97
	Yichun	27.81N 114.42E	152.5	105.75
	Pingxiang	27.62N 113.85E	112.43	123.82
	Zhuzhou	27.83N 113.13E	48.86	158.82
	Xiangtan	27.83N 112.94E	44.99	179.57
	Hengyang	26.89N 112.57E	153.68	195.26
	Shaoyang	27.25N 111.47E	181.64	236.32
	Loudi	27.69N 111.99E	110.36	240.4
	Yiyang	28.55N 112.36E	66.7	299.25
	Changde	29.03N 111.7E	149.44	302.83
Chongqing				
29.56N 106.54E	Guangan	30.45N 106.64E	99.19	5.91
	Dazhou	31.21N 107.47E	203.96	29.35
	Zunyi	27.73N 106.92E	207.33	168.36
	Luzhou	28.86N 105.44E	133.13	237.67
	Yibin	28.73N 104.65E	206.64	246.43
	Zigong	29.33N 104.78E	173.39	262.6
	Neijiang	29.58N 105.05E	145.03	270.69
	Ziyang	30.13N 104.63E	195.25	286.45
	Suining	30.54N 105.59E	142.07	315.65
	Nanchong	30.85N 106.13E	148.98	342.35

continued

Target Cities	Instrumental Cities	Coordinate (°)	Distance (km)	Location (°)
Guangzhou	Shaoguan	24.8N 113.6E	118.05	11.33
23.12N 113.27E	Heyuan	23.76N 114.7E	162.57	65.87
	Huizhou	23.11N 114.41E	117.49	90.5
	Dongguan	23.02N 113.75E	52.29	101.49
	Shenzhen	22.55N 114.06E	104.1	125.91
	Zhuhai	22.28N 113.58E	99.92	159.772
	Jiangmen	22.58N 113.08E	64.29	119.03
	Yangjiang	21.86N 111.98E	192.95	225.52
	Foshan	23.03N 113.13E	18.27	236.22
	Yunfu	22.91N 112.04E	127.79	260.39
	Zhaoqing	23.02N 112.48E	80.5	262.94
	Qingyuan	23.68N 113.06E	64.85	339.75
Hangzhou				
30.28N 120.15E	Suzhou	31.32N 120.59E	118.59	22.83
	Jiaxing	30.75N 120.76E	78.84	52.18
	Shanghai	31.23N 121.47E	164.45	53.97
	Zhoushan	30.02N 122.21E	200.09	97.1
	Ningbo	29.88N 121.54E	142.69	105.28
	Shaoxing	30.01N 120.61E	52.9	124.5
	Lishui	28.48N 119.95E	203.64	186.61
	Jinhua	29.06N 119.65E	140.28	202.57
	Quzhou	29N 118.9E	188.8	224.55
	Huangshan	29.72N 118.38E	183.42	252.56
	Xuancheng	30.94N 118.76E	152.66	295.32
	Wuhu	31.37N 118.42E	203.74	302.2
	Huzhou	30.89N 120.08E	68.97	355.07
	Changzhou	31.81N 119.97E	172.53	353.28

continued

Target Cities	Instrumental Cities	Coordinate (°)	Distance (km)	Location (°)
Nanjing	Huaian	33.6N 119.02E	172.28	8.4
32.06N 118.79E	Yangzhou	32.38N 119.41E	68.42	62.56
	Taizhou	32.45N 119.91E	114.29	70.73
	Zhenjiang	32.2N 119.43E	61.71	77.84
	Nantong	31.96N 120.89E	199.18	92.65
	Changzhou	31.81N 119.97E	115.35	102.144
	Wuxi	31.48N 120.3E	156.81	110.829
	Suzhou	31.32N 120.59E	189.96	112.88
	Huzhou	30.89N 120.08E	178.78	132.28
	Xuancheng	30.94N 118.76E	126.48	181.46
	Wuhu	31.37N 118.42E	86.46	206.08
	Maanshan	31.66N 118.51E	51.54	215.29
	Tongling	30.94N 177.82E	153.9	220.96
	Chizhou	30.66N 117.5E	198.53	222.71
	Hefei	31.83N 117.23E	149.82	261.52
	Huainan	32.64N 117.01E	179.49	288.12
	Chuzhou	32.25N 118.33E	48.49	292.57
	Bengbu	32.91N 117.39E	162.03	301.4
	Suqian	33.96N 118.28E	216.55	345.06
Shanghai	Zhoushan	30.02N 122.21E	155.64	149.15
31.23N 121.47E	Ningbo	29.88N 121.54E	150.87	177.17
	Shaoxing	30.01N 120.61E	157.34	215.32
	Hangzhou	30.28N 120.15E	164.45	233.37
	Suzhou	31.32N 120.59E	82.2	275.66
	Changzhou	31.81N 119.97E	154.08	291.72
	Taizhou	32.45N 119.91E	198.03	308.74
	Nantong	31.96N 120.89E	100.1	322.69

continued

Target Cities	Instrumental Cities	Coordinate (°)	Distance (km)	Location (°)
Tianjin	Tangshan	39.62N 118.18E	103.88	60.84
39.09N 117.19E	Binzhou	37.39N 117.97E	200.5	155.381
	Cangzhou	38.31N 116.84E	92.16	304.27
	Dezhou	37.44N 116.36E	198.12	206.83
	Hengshui	37.73N 115.66E	200.69	228.52
	Baoding	38.89N 115.47E	151.83	263.06
	Langfang	39.54N 116.68E	65.5	310.95
	Beijing	39.96N 116.43E	113.34	319.13
Wuhan	Huanggang	30.45N 14.88E	56.93	104.84
30.61N 114.33E	Hangshi	30.19N 115.05E	84.53	119.04
	Jiujiang	29.71N 116E	191.59	117.81
	Xianning	29.83N 114.33E	84.72	178.51
	Yueyang	29.36N 113.31E	178.8	222.77
	Jinzhou	30.34N 112.24E	198.96	262.81
	Xiaogan	30.91N 113.94E	49.14	310.22
	Suizhou	31.69N 113.4E	149.01	319.99
	Xinyang	32.15N 114.09E	173.69	351.97
Zhengzhou	Xinxiang	35.3N 113.93E	66.93	29.87
34.75N 113.62E	Hebi	35.75N 114.3E	126.91	34.16
	Anyang	36.1N 114.4E	165.73	29.97
	Puyang	35.76N 115.02E	169.63	54.33
	Heze	35.23N 115.49E	178.36	75.53
	Kaifeng	34.78N 114.31E	63.32	87.35
	Shangqiu	34.43N 115.65E	189.54	99.09
	Zhoukou	33.62N 114.69E	159.6	136.64
	Xuchang	34.04N 113.84E	81.39	162.59
	Pingdingshan	33.76N 113.84E	116.39	202.83
	Luoyang	34.6N 112.46E	107.12	262.81
	Jiaozuo	35.21N 113.24E	61.4	320.32
	Changzhi	36.19N 113.12E	166.53	341.03

Table A9: Robustness — Accounting for Daily Wind Speed in IV

	AQI		PM2.5	
	(1)	(2)	(3)	(4)
First Stage[(a)]				
$P_{source_{it}}$	0.650***	0.280***	0.668***	0.289***
	(0.054)	(0.064)	(0.047)	(0.078)
$P_{source_{it}} * windspeed_{it}$	-0.004	0.003	-0.012	-0.001
	(0.008)	(0.008)	(0.008)	(0.009)
$P_{source_{it-1}}$		0.699***		0.738***
		(0.073)		(0.086)
$P_{source_{it}} * windspeed_{it}$		-0.040***		-0.050***
		(0.006)		(0.007)
F-statistic	162.64	372.80	200.19	244.24
Second Stage[(b)]				
Instrumented pollutant	0.198**	0.164***	0.222*	0.177**
	(0.089)	(0.059)	(0.115)	(0.073)
Additional Control				
City FEs	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y

Notes: (a) Dependent variable in the first stage is daily-mean pollutant of target city, and independent variable is source pollutant ($100km < d_{ij} < 300km$) from upwind direction (within 90 degrees to the wind), and its interaction term with wind speed in the target city. (b) Second stage reports the results regressing log form of Sleeplessness Index on the instrumented daily pollution. Column (2) and (4) incorporate day before as additional instrument. Temporal controls include year_season, day of week and holiday fixed effects. Weather controls contain temperature, humidity, precipitation, wind speed and sea-level pressure. Temperature and humidity are measured by the way of bins. Pairs bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A10: Robustness — 100 km to 300 km Inclusion Criterion in IV

2SLS	AQI		PM2.5	
	(1)	(2)	(3)	(4)
First Stage[(a)]				
Instrumental Pollutant t	0.644***	0.271***	0.615***	0.260***
	(0.075)	(0.066)	(0.083)	(0.067)
Instrumental Pollutant lagged t-1		0.549***		0.517***
		(0.084)		(0.083)
F-statistics	74.13	144.49	55.07	90.04
Second Stage[(b)]				
Instrumented Pollutant	0.208***	0.195***	0.242**	0.229***
	(0.082)	(0.070)	(0.100)	(0.088)
Observations	6573	6526	6573	6526
Additional Control				
City FEs	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y

Notes: (a) Dependent variable in the first stage is daily-mean pollutant of target city, and independent variable is daily weighted average pollution of surrounding cities ($100km < d_{ij} < 300km$) from upwind direction (within 90 degrees to the wind). (b) Second stage reports the results regressing Sleeplessness Index on the instrumented daily pollution. Column (2) and (4) incorporate day before as additional instrument. Temporal controls include year_season, day of week and holiday fixed effects. Weather controls contain temperature, humidity, precipitation, wind speed and sea-level pressure. Temperature and humidity are measured by the way of bins. Pairs bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A11: City Sub-samples — OLS

	Full	Exclude Beijing and Environ	Exclude Shanghai and Environs
	(1)	(2)	(3)
Panel A: AQI	0.043***	0.068***	0.052***
	(0.014)	(0.012)	(0.015)
Panel B: PM2.5	0.049***	0.077***	0.060***
	(0.017)	(0.013)	(0.017)
Number of Observations	6839	5478	4770
Additional Controls			
City FEs	Y	Y	Y
Temporal Controls	Y	Y	Y
Weather Covariates	Y	Y	Y

Notes: Column (1) replicates the OLS results in Column (5) of Table 4. Column (2) excludes Beijing and its nearby city, Tianjin, both of which are situated in northern heavy industrial region. Column (3) excludes Shanghai and its nearby cities, Nanjing and Hangzhou, which are coastally located and dominated by light industry. All the regressions include city fixed effects, temporal controls and weather controls. Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A12: City Sub-samples — IV

	Full	Exclude Beijing and Environ	Exclude Shanghai and Environs
	(1)	(2)	(3)
		Panel A: AQI	
First Stage			
Instrumental AQI t	0.277***	0.314***	0.279***
	(0.035)	(0.045)	(0.046)
Instrumental AQI lagged t-1	0.489***	0.501***	0.421***
	(0.062)	(0.072)	(0.053)
F-Statistic	184.7	510.9	145.47
Second Stage			
Instrumented Pollutant	0.196**	0.289***	0.215**
	(0.084)	(0.059)	(0.098)
		Panel B: PM2.5	
First Stage			
Instrumental $PM_{2.5}$ t	0.258***	0.261***	0.255***
	(0.045)	(0.056)	(0.058)
Instrumental $PM_{2.5}$ lagged t-1	0.463***	0.540***	0.384***
	(0.068)	(0.079)	(0.068)
F-Statistic	108.94	273.29	88.11
Second Stage			
Instrumented Pollutant	0.224**	0.326***	0.255*
	(0.110)	(0.072)	(0.136)
Number of Observations	6475	5118	4553
Additional Controls			
City FEs	Y	Y	Y
Temporal Controls	Y	Y	Y
Weather Covariates	Y	Y	Y

Notes: Column (1) replicates the IV results in Column (6) of Table 5. Column (2) excludes Beijing and its nearby city, Tianjin, both of which are situated in northern heavy industrial region. Column (3) excludes Shanghai and its nearby cities, Nanjing and Hangzhou, which are coastally located and dominated by light industry. All the regressions include city fixed effects, temporal controls and weather controls. Temperature and humidity are measured in the form of bins. Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A13: Individual City Effect — AQI

Independent Variable (Daily Pollutant)	Dependent Variable (Ln(Sleepless))										
Individual City	Full	Beijing	Changsha	Chongqing	Guangzhou	Hangzhou	Nanjing	Shanghai	Tianjin	Wuhan	Zhengzhou
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
						Panel A1: AQI-OLS					
AQI	0.043***	0.032**	0.058***	0.088***	0.185***	0.039	0.070***	0.008	0.064***	0.107***	0.047**
(OLS)	(0.014)	(0.016)	(0.015)	(0.033)	(0.055)	(0.032)	(0.023)	(0.030)	(0.017)	(0.024)	(0.023)
Observations	6839	665	730	646	660	681	674	714	696	694	679
						Panel A2: AQI-IV					
First Stage											
Instrumental AQI t	0.277***	0.198*	0.246	0.765***	0.564***	0.143	0.447**	0.238**	0.118	0.295**	0.391
	(0.035)	(0.116)	(0.213)	(0.246)	(0.202)	(0.180)	(0.201)	(0.116)	(0.166)	(0.126)	(0.321)
Instrumental AQI lagged t-1	0.489***	0.312***	0.626***	0.128	0.174	0.499***	0.688***	0.703***	0.684***	0.571***	0.156
	(0.062)	(0.106)	(0.217)	(0.243)	(0.197)	(0.182)	(0.201)	(0.115)	(0.164)	(0.112)	(0.306)
F-statistics	184.7	8.31	21.46	42.96	26.09	6.62	26.21	30.62	16.47	39.37	13.16
Second Stage											
Instrumented AQI	0.196**	0.186	0.195***	0.546***	0.996***	0.917***	0.123**	0.011	-0.152**	0.216***	0.395***
	(0.084)	(0.118)	(0.055)	(0.010)	(0.177)	(0.269)	(0.051)	(0.086)	(0.078)	(0.050)	(0.104)
Observations	6475	662	530	645	659	563	673	686	695	684	678
Additional Control											
City FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Column (1) repeats the preferred results for AQI in the last column of Table (4) and Table (5). Column (2) through Column (11) present individual health effect of AQI for each city via both OLS and IV. The instruments are as used in the Column (6) of Table 5. All the regressions include city fixed effects, temporal controls and weather controls. Temperature and humidity are measured in the form of bins. Robust standard errors are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A14: Individual City Effect — PM2.5

Independent Variable (Daily Pollutant)	Dependent Variable (Ln(Sleepless))										
Individual City	Full	Beijing	Changsha	Chongqing	Guangzhou	Hangzhou	Nanjing	Shanghai	Tianjin	Wuhan	Zhengzhou
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					Panel B1: PM2.5-OLS						
PM2.5	0.049***	0.033*	0.064***	0.110***	0.239***	0.049	0.078***	0.009	0.077***	0.096***	0.065***
(OLS)	(0.017)	(0.020)	(0.018)	(0.040)	(0.069)	(0.039)	(0.027)	(0.036)	(0.021)	(0.029)	(0.027)
Observations	6839	665	730	646	660	681	674	714	696	694	679
					Panel B2: PM2.5-IV						
First Stage											
Instrumental PM2.5 t	0.258***	0.190*	0.156	0.749***	0.628***	0.139	0.407**	0.290***	0.128	0.165	0.428
	(0.045)	(0.113)	(0.209)	(0.237)	(0.193)	(0.170)	(0.207)	(0.107)	(0.164)	(0.205)	(0.317)
Instrumental PM2.5 lagged t-1	0.463***	0.248***	0.679***	0.105	0.146	0.502***	0.756***	0.663***	0.517***	0.717***	0.123
	(0.068)	(0.100)	(0.209)	(0.234)	(0.185)	(0.169)	(0.205)	(0.108)	(0.166)	(0.171)	(0.306)
F-statistics	108.94	7.31	21.05	42.15	30.73	7.63	23.77	34.25	11.83	39.14	14.49
Second Stage											
Instrumented PM2.5	0.224**	0.202	0.251***	0.685***	1.034***	1.012***	0.136**	-0.024	-0.304***	0.223***	0.434***
	(0.110)	(0.153)	(0.068)	(0.125)	(0.187)	(0.278)	(0.059)	(0.094)	(0.121)	(0.059)	(0.113)
Observations	6475	662	530	645	659	563	673	686	695	684	678
Additional Control											
City FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Column (1) repeats the preferred results for PM2.5 in the last column of Table (4) and Table (5). Column (2) through Column (11) present individual health effect of PM2.5 for each city via both OLS and IV. The instruments are as used in the Column (6) of Table 5. All the regressions include city fixed effects, temporal controls and weather controls. Temperature and humidity are measured in the form of bins. Robust standard errors are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A15: Alternative Clusters — OLS

Alternative Clusters (OLS)	City (1)	City_year_season (2)	City_year_month (3)	City_year_week (4)
Panel A: AQI	0.043***	0.043***	0.043***	0.043***
	(0.014)	(0.013)	(0.014)	(0.012)
Panel B: PM2.5	0.049***	0.049***	0.049***	0.049***
	(0.017)	(0.014)	(0.017)	(0.013)
Clusters	10	80	236	1040
Observations	6839	6839	6839	6839
Additional Control				
City FEs	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y

Notes: Dependent variable is log form of Sleeplessness Index. Independent variable is city daily-mean value of specific pollutant. Column (1) displays the results using pairs bootstrapped standard errors clustered at city level which replicates the results under Column (5) in Table 4. Column (2) through Column (4) are clustered at city_year_season, city_year_month and city_year_week, respectively. All the regressions include city fixed effects, temporal controls and weather covariates. Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A16: Alternative Clusters — IV

Alternative Clusters (2SLS)	City (1)	City_year_season (2)	City_year_month (3)	City_year_week (4)
		Panel A: AQI		
First Stage				
Instrumental AQI t	0.277***	0.277***	0.277***	0.277***
	(0.035)	(0.059)	(0.061)	(0.051)
Instrumental AQI lagged t-1	0.489***	0.489***	0.489***	0.489***
	(0.062)	(0.064)	(0.051)	(0.052)
F-statistics	184.7	40.59	65.74	97.06
Second Stage				
Instrumented AQI	0.196**	0.196**	0.196**	0.196***
	(0.084)	(0.085)	(0.084)	(0.049)
		Panel B: PM2.5		
First Stage				
Instrumental PM2.5 t	0.258***	0.258***	0.258***	0.258***
	(0.045)	(0.060)	(0.072)	(0.062)
Instrumental PM2.5 lagged t-1	0.463***	0.463***	0.463***	0.463***
	(0.068)	(0.078)	(0.061)	(0.059)
F-statistics	108.94	37.05	55.44	82.51
Second Stage				
Instrumented PM2.5	0.224**	0.224**	0.224**	0.224***
	(0.110)	(0.108)	(0.107)	(0.061)
Clusters	10	80	236	1040
Observations	6475	6475	6475	6475
Additional Control				
City FEs	Y	Y	Y	Y
Temporal Controls	Y	Y	Y	Y
Weather Covariates	Y	Y	Y	Y

Notes: Dependent variable in the first stage is daily-mean pollutant of local city, and independent variable is daily weighted average pollution of source cities. Second stage reports the results regressing log Sleeplessness Index on the instrumented daily pollution. Column (1) repeats the IV results of Column (6) in Table 5, in which pairs bootstrapped standard errors clustered at city level. Column (2) through Column (4) are clustered at alternative clusters indicated on the first row. All the regressions include city fixed effects, temporal controls and weather covariates. Temperature and humidity are measured in the form of bins. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A17: Air Quality and Sleeplessness — OLS All Coefficients

	AQI		PM2.5	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Pollutant	0.043***	0.196**	0.049***	0.224**
	(0.014)	(0.084)	(0.017)	(0.110)
Average Temperature (T ∈ [10, 15) Omitted)				
T ∈ [-5, 0)	14.678***	12.140***	14.734***	12.458***
	(2.579)	(3.554)	(2.563)	(3.448)
T ∈ [0, 5)	8.376***	5.713***	8.386***	5.754***
	(1.229)	(1.854)	(1.200)	(1.901)
T ∈ [5, 10)	3.748*	1.039	3.797*	1.259
	(2.137)	(2.145)	(2.122)	(2.200)
T ∈ [15, 20)	-0.169	2.099	-0.210	1.922
	(1.949)	(1.899)	(1.963)	(1.906)
T ∈ [20, 25)	2.922	6.751**	2.783	6.146**
	(3.262)	(2.861)	(3.286)	(2.908)
T ∈ [20, 25)	-0.152	2.552	-0.287	1.927
	(3.941)	(3.422)	(3.974)	(3.503)
T ≥ 30	-12.835*	-11.925*	-12.919*	-12.195*
	(7.874)	(6.948)	(7.910)	(6.992)
Max Temperature	-0.541**	-0.883***	-0.530**	-0.843***
	(0.246)	(0.269)	(0.247)	(0.284)
Min Temperature	-0.411*	-0.182	-0.401*	-0.122
	(0.237)	(0.237)	(0.236)	(0.255)
Precipitation	0.047	0.049	0.047	0.049
	(0.044)	(0.038)	(0.044)	(0.040)
Sea-level Pressure	-0.887***	-0.711***	-0.885***	-0.702***
	(0.123)	(0.084)	(0.120)	(0.086)
Wind Speed	-0.241	0.071	-0.234	0.101
	(0.206)	(0.188)	(0.204)	(0.208)

Continued

	AQI		PM2.5	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Average Humidity (H [40,60) Omitted)				
H <20	-0.891	-1.429	-1.194	-2.738
	(5.033)	(4.413)	(4.925)	(4.721)
H ∈ [20, 40)	-3.629	-9.880	-4.079	-11.929
	(6.692)	(7.292)	(6.502)	(8.157)
H ∈ [60, 80)	-2.583	-10.816	-3.095	-13.226
	(6.734)	(7.261)	(6.510)	(8.463)
H ≥ 80	-1.754	-9.498	-2.276	-11.932
	(7.401)	(7.164)	(7.158)	(8.157)
Max Humidity	-0.071	-0.089	-0.073	-0.097
	(0.087)	(0.100)	(0.087)	(0.103)
Min Humidity	0.004	0.012	-0.002	-0.017
	(0.056)	(0.059)	(0.056)	(0.056)
Day of Week	(Monday Omitted)			
Tuesday	0.214	0.053	0.208	0.044
	(0.290)	(0.412)	(0.288)	(0.431)
Wednesday	-1.555***	-1.587***	-1.586***	-1.722***
	(0.553)	(0.610)	(0.557)	(0.589)
Thursday	-0.983*	-0.756	-0.997*	-0.807
	(0.600)	(0.705)	(0.597)	(0.687)
Friday	-3.781***	-3.755***	-3.800***	-3.819***
	(0.748)	(0.791)	(0.747)	(0.783)
Saturday	-0.197	-0.648	-0.289	-1.122
	(0.993)	(1.146)	(0.967)	(1.289)
Sunday	12.333***	11.480***	12.295***	11.267***
	(1.084)	(1.097)	(1.070)	(1.183)
Holiday	1.181	1.491	1.251	1.880
	(1.590)	(1.463)	(1.574)	(1.417)

Continued

	AQI		PM2.5	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
2014_Spring (Omitted)				
2014_Summer	-19.246***	-13.210***	-19.234***	-13.078***
	(2.608)	(3.602)	(2.600)	(3.881)
2014_Autumn	-9.702*	-7.006	-9.713*	-7.042
	(5.169)	(5.836)	(5.180)	(5.766)
2014_Winter	14.982***	13.832***	15.024***	13.980***
	(5.093)	(5.208)	(5.128)	(5.168)
2015_Spring	46.380***	50.247***	46.463***	50.646***
	(7.818)	(6.721)	(7.835)	(6.771)
2015_Summer	5.988	15.472***	5.937	15.294***
	(6.836)	(4.891)	(6.838)	(4.930)
2015_Autumn	-28.533***	-22.964***	-28.605***	-23.231***
	(3.424)	(4.315)	(3.439)	(4.305)
2015_Winter	5.174	2.528	5.277	2.928
	(4.968)	(5.623)	(4.980)	(5.535)

Notes: The table reports detailed OLS and IV results. Each column represents a separate regression. Dependent variable is log form of Sleeplessness Index. Independent variables include daily mean level of specific pollutant, weather controls (average temperature bins, max and min temperature, precipitation, sea-level pressure, wind speed, average humidity, max and min humidity), temporal controls (year_season fixed effects, day of week dummies and holiday dummy) and city fixed effects. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).

Table A18: Pollution Regressed on Imported Source Pollutants All Coefficients (First Stage)

First Stage	**AQI** **(1)**	$PM_{2.5}$ (2)
Instrumental	0.277***	0.258***
Pollutant t	(0.035)	(0.045)
Instrumental	0.489***	0.463***
Pollutant lagged t-1	(0.062)	(0.068)
Average Temperature (T $\in$ [10, 15) Omitted)		
T $\in$ [-5, 0)	-6.686	-9.421
	(13.194)	(11.569)
T $\in$ [0, 5)	-1.677	-1.845
	(7.966)	(7.048)
T $\in$ [5, 10)	8.065**	6.744*
	(4.151)	(3.747)
T $\in$ [15, 20)	-9.472***	-7.503***
	(2.828)	(2.364)
T $\in$ [20, 25)	-19.212***	-14.162***
	(5.972)	(5.295)
T $\in$ [25, 30)	-13.517*	-9.615
	(7.394)	(6.925)
T $\geq$ 30	-14.660	-12.608
	(9.215)	(8.578)
Max Temperature	2.722***	2.208***
	(0.670)	(0.683)
Min Temperature	-1.452***	-1.567***
	(0.460)	(0.476)
Precipitation	-0.105	-0.087
	(0.069)	(0.056)
Sea-level Pressure	-0.476*	-0.509
	(0.292)	(0.221)
Wind Speed	-1.040***	-1.043***
	(0.210)	(0.200)
		Continued

First Stage	**AQI** **(1)**	$PM_{2.5}$ (2)
Average Humidity (H ∈ [40,60) Omitted)		
H <20	3.150	8.579**
	(8.668)	(4.042)
H ∈ [20, 40)	29.677*	34.107***
	(16.184)	(9.295)
H ∈ [60, 80)	40.280**	45.213***
	(19.162)	(12.615)
H ≥ 80	35.049**	41.038***
	(18.000)	(12.042)
Max Humidity	0.227***	0.201***
	(0.085)	(0.071)
Min Humidity	0.095	0.183
	(0.156)	(0.144)
Day of Week	(Monday Omitted)	
Tuesday	-2.096**	-1.796**
	(1.002)	(0.879)
Wednesday	-2.977**	-2.058*
	(1.419)	(1.129)
Thursday	-2.237*	-1.692
	(1.289)	(1.164)
Friday	-1.725	-1.193
	(1.110)	(90.887)
Saturday	5.597***	7.012***
	(1.630)	(1.477)
Sunday	7.903***	7.828***
	(1.707)	(1.637)
Holiday	-6.823***	-7.622***
	(1.663)	(1.485)
		Continued

First Stage	**AQI** **(1)**	$PM_{2.5}$ (2)
2014_Spring (Omitted)		
2014_Summer	-16.746****	-16.988***
	(5.376)	(5.215)
2014_Autumn	-1.273	-2.991
	(5.447)	(4.589)
2014_Winter	17.117****	13.310***
	(4.435)	(3.768)
2015_Spring	-0.802	-3.552
	(3.759)	(3.003)
2015_Summer	-20.643***	-20.580***
	(6.320)	(5.988)
2015_Autumn	-5.617*	-7.981**
	(3.344)	(3.328)
2015_Winter	20.433***	14.468***
	(6.268)	(5.122)

Notes: The table reports detailed results of the first stage under IV estimations. Each column represents a separate regression. Dependent variable is the daily mean level of the specific pollutant for each city. Independent variables include imported pollutants from instrumental cities, weather controls (average temperature bins, max and min temperature, precipitation, sea-level pressure, wind speed, average humidity, max and min humidity), temporal controls (year_season fixed effects, day of week dummies and holiday dummy) and city fixed effects. Bootstrapped standard errors clustered at city level are reported in parentheses (* significant at 10%, ** significant at 5%, *** significant at 1%).