Journal of Environmental Economics and Management 119 (2023) 102810

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

# Flight delays due to air pollution in China

Xiaoguang Chen [a,1], Luoye Chen [b,1], Wei Xie [c,*,1], Nathaniel D. Mueller [d,e], Steven J. Davis [f,g]

[a] Research Institute of Economics and Management, Southwestern University of Finance and Economics, Chengdu, 610074, China
[b] Innovation, Policy, and Entrepreneurship Thrust, Society Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou, 511458, China
[c] China Center for Agricultural Policy, School of Advanced Agricultural Sciences, Peking University, Beijing, 100871, China
[d] Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, 80523, USA
[e] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, 80523, USA
[f] Department of Earth System Science, University of California, Irvine, CA, 92697, USA
[g] Department of Civil and Environmental Engineering, University of California, Irvine, CA, 92697, USA

ARTICLE INFO

*JEL classification:*
L93
Q53
Q58

*Keywords:*
$PM_{2.5}$ pollution
Flight delays
Clean air policy
Economic benefits
China

ABSTRACT

Using a newly-compiled dataset of airport traffic, air pollution, and weather from 2015 to 2017, we evaluate the effect of $PM_{2.5}$ pollution levels on flight departure delays in China. We instrument for $PM_{2.5}$ levels using changes in atmospheric thermal inversions and wind direction. We find that a one-standard-deviation increase in the daily $PM_{2.5}$ concentration levels leads to roughly 4.5–6.4 additional minutes in departure delays per flight. We estimate that reductions in China's $PM_{2.5}$ pollution levels due to the Air Pollution Prevention and Control Action Plan avoided 13–18% of the increases in national average flight delays and saved US $54 to $69 million in passenger time in 2017 alone.

## 1. Introduction

Airport delays cause substantial economic losses to millions of passengers, airlines, and the global economy (Ball et al., 2010). In the U.S. and Europe, such delays have been attributed primarily to increases in market concentration, airport congestion, and weather shocks (Deshpande and Arıkan, 2012; Mayer and Sinai, 2003; Santos and Robin, 2010). However, recent studies have suggested that severe air pollution may be an important driver of flight delays in developing countries, along with inefficient airport operations and outdated traffic control procedures (AlKheder and AlKandari, 2020; Baddock et al., 2013). Although research has linked air pollution to adverse health consequences (Chay and Greenstone, 2003; Chen et al., 2013; Lelieveld et al., 2015), impaired cognitive performance (Carneiro et al., 2021; Ebenstein et al., 2016; Zhang et al., 2018), reduced labor supply (Hanna and Oliva, 2015), lower work productivity (He et al., 2019; Zivin and Neidell, 2012), decreased happiness (Zhang et al., 2017; Zheng et al., 2019), and road traffic accidents (Sager, 2019), the relationship between air pollution and flight delays or cancellations has not been quantitatively assessed.[2]

* Corresponding address.
*E-mail address:* xiewei.ccap@pku.edu.cn (W. Xie).

[1] X.C., L.C., and W.X. contributed equally to this work and jointly share the first authorship.

[2] Several media sources have reported that air pollution caused flight delays. For instance, see "Smog in northern China shuts highways, causes cancelled flights," *Reuters*, January 1, 2017.

https://doi.org/10.1016/j.jeem.2023.102810
Received 21 April 2021; Received in revised form 28 February 2023; Accepted 1 March 2023
Available online 4 March 2023
0095-0696/

In this paper, using a newly-compiled dataset of airport traffic, air pollution, and weather from 2015 to 2017, we evaluate the impacts of air pollution on flight departure delays in China. There are several reasons for this focus. First, China has become the world's largest aviation market, with over 610 million passengers carried every year since 2017 (The World Bank, 2018). Second, flight delays have become a pressing issue in China. According to the Civil Aviation Administration of China (CAAC), approximately 8% of the total departures (~243 thousand flights) were delayed for more than 2 h in 2017 in large Chinese airports that transport over 10 million passengers each year. Flight delays routinely account for about 50% of all customer complaints to the CAAC (). Finally, despite significant progress in reducing pollution emissions (Zhang et al., 2019), major Chinese cities such as Beijing and Shanghai still rank among the most polluted in the world. Hundreds of flights were either severely delayed or canceled daily in the winters of 2012, 2013, and 2016, as air pollution shrouded these two cities (Tang and Hoshiko, 2013; Zhuang, 2016).

Air pollution may cause flight delays or cancellations for two possible reasons. First, particulate matter can reduce atmospheric visibility (Hyslop, 2009). Poor visibility at airports can downgrade the capacity of an airport and therefore lead to delays or cancellations of departing flights. Second, air pollution reduces the cognitive performance of airport ground crew and airline cabin crew, which may increase the possibility of flights being delayed or cancelled. Several studies document negative correlations between air pollution and cognitive performance (Carneiro et al., 2021; Ebenstein et al., 2016; Zhang et al., 2018). Ebenstein et al. (2016) find that transitory $PM_{2.5}$ exposure is associated with a significant decline in student academic performance. Thus, short-term exposure to $PM_{2.5}$ is expected to negatively affect cognitive performance of airport ground and cabin crews. Because efficient air transportation is a complex task that requires tight coordination among airport ground and cabin crews, the efficiency of airport operations could be significantly reduced by air pollution. However, the literature has so far overlooked the potential impact of air pollution on flight delays.[3]

In this study, we first investigate the impact of air pollution, measured as the daily average concentrations of $PM_{2.5}$ on flight departure delays in China between 2015 and 2017. Second, we assess the extent to which improved air quality induced by more stringent air pollution policies contributed to the change in flight delays in China from 2015 to 2017. China has been undertaking aggressive actions to combat air pollution since the promulgation of the "Air Pollution Prevention and Control Action Plan" in 2013. Significant declines in $PM_{2.5}$ concentrations have been observed nationwide from 2013 to 2017. (Zhang et al., 2019). The answer to our second research question provides critical information about the potential gains from improving air quality in China.

The dataset used in our analysis contains flight-level information for all domestic and international air carriers landing at or departing from airports in North China (including Beijing, Tianjin, Hebei, Inner Mongolia and Shanxi provinces) from April 1, 2015 to September 30, 2017. We use two methods to measure flight departure delays. We first calculate a flight's *gross* delay based on its actual and scheduled departure times. For flights with previous assignments, we deduct these flights' arrival delays from the gross delays to calculate *net* flight delays in the current assignment. Because the raw high-frequency flight-by-flight dataset has a considerable number of missing observations, following Schlenker and Walker (2016), we aggregate the *gross* (or *net*) delays of individual flights to obtain daily total flight delays at the airport level; we then measure flight delays using daily average delay per flight (= daily total flight delays/daily total number of flights).

Estimation of the impacts of pollution on flight delays is challenging because airports constitute a major source of air pollutants. Airport runway congestion due to bad weather, network delays, or air traffic controls contributes significantly to local air pollution (Schlenker and Walker, 2016). To credibly identify the causal effects of $PM_{2.5}$ on flight departure delays, we use two instrumental variable (IV) strategies, where we rely on changes in atmospheric thermal inversions and wind direction as exogenous shocks to local pollution levels. Prior studies have used thermal inversions as an instrument for $PM_{2.5}$ when estimating the effects of $PM_{2.5}$ on a wide range of outcomes, such as health (Arceo et al., 2016; Jans et al., 2018), productivity (Fu et al., 2021; He et al., 2019), road safety (Sager, 2019) and crime (Bondy et al., 2020). Several recent studies also show that wind direction induces an exogenous shock to local pollution concentrations and can be used as an instrumental variable for $PM_{2.5}$ (Bondy et al., 2020; Carneiro et al., 2021; Deryugina et al., 2019).

Our results suggest that exposure to rising levels of $PM_{2.5}$ significantly increases flight departure delays. We find that, on average, a one-standard-deviation increase in the daily $PM_{2.5}$ concentration leads to 6.4 additional minutes in gross departure delay per flight and 4.5 additional minutes in net departure delay per flight after excluding arrival delays from previous assignments, with $PM_{2.5}$ instrumented by wind direction. The estimated impacts on flight delays remain statistically similar when instrumenting pollution with thermal inversions. These IV estimates are robust to a wide range of variations in specifications, data, and estimation strategies.

A key identifying assumption of our IV approaches is that, after controlling for weather conditions and relevant fixed effects, changes in thermal inversions and wind direction are not related to changes in flight departure delays, except through their influences on pollution levels. This assumption could be violated because both instruments are weather phenomena, and they may be correlated with unobserved weather events that affect flight planning. To ensure robustness of our findings, we augment our main specification by adding flexible weather controls. We also analyze subsamples that exclude days with inclement weather for flight operations. It is reassuring that our results are robust to these additional analyses, thus providing strong supportive evidence of the robustness of our estimates to the presence of unobserved weather confounding factors.

We also investigate the validity of the monotonicity assumption when using wind direction as an instrument; this assumption is less of a concern for the inversion instrument because it is unlikely that the pollution-inversion relationship holds in some regions but is

[3] Several studies have attempted to explain the reasons for flight delays. For instance, Mayer and Sinai (2003) investigated the effect of airline hub size and airport concentration on air travel time. Rupp and Holmes (2006) examined several determinants of flight cancellations, including competition and revenue.

reversed in other regions. More specifically, we estimate alternative specifications by allowing instruments to vary with the size of wind direction bins and the number of groups of air pollution monitors (as explained below). Our results remain robust to these variations, which lends support to the monotonicity assumption for the wind direction instrument. Our IV estimates are also robust to simultaneously instrumenting for $PM_{2.5}$ and other air pollutants, including sulfur dioxide, carbon monoxide, ozone, and nitrogen dioxide, suggesting that the estimated impacts of air pollution on flight delays are indeed due to $PM_{2.5}$ and are not related to other air pollutants.

These adverse effects of pollution on flight delays are more pronounced for passengers travelling long distances and departing from large airports. The observed national average *net* (*gross*) delay per flight increased by 4.6 (4.5) minutes in 2017 compared to that in 2015. We estimate that the reduction in $PM_{2.5}$ concentrations attributable to clean air policies in China, avoided approximately 13–18% of the increases in national average flight delays and saved US $54 to $69 million in passenger time in 2017 alone.

Our findings have a number of implications. First, the robust negative correlation between $PM_{2.5}$ and flight departure delays demonstrates a substantial impact of air pollution that has been neglected by current estimates of social costs of air pollution, which primarily focus on health outcomes and productivity. Second, our results suggest that improving air quality at airports presents an opportunity to reduce flight departure delays, especially for flights departing from large airports and travelling long distances on polluted days. Finally, our results further support China's ambitious clean air policies by showing that the improvement in air quality in China has led to a considerable economic benefit due to avoided flight delays.

The rest of the paper is organized as follows. Section 2 describes data. Section 3 presents our empirical strategy, and Section 4 reports our results. We conclude in Section 5.

## 2. Data

We compile a comprehensive dataset on airport traffic, air pollution, weather and thermal inversions in China between 2015 and 2017. This dataset is rich in both temporal and spatial dimensions, allowing for a fine-grained analysis of how air pollution affects flight delays. This section reports data sources and presents descriptive statistics.

### 2.1. Flight delays

We obtain flight information from the North China Regional Administration of the Air Traffic Management Bureau of CAAC. The unique dataset contains flight-level information by all certified domestic and international air carriers departing from or arriving at airports in North China from April 1, 2015 to September 30, 2017. The dataset includes flight number, aircraft type, departure and arrival airports, scheduled and actual departure and arrival times, taxiing time, and the previous flight and next flight assigned to the airplane. Our sample covers 89 airports in China (Fig. 1 and Appendix Fig. S1). Similarly to the flight-level dataset in a case study in the United States (Schlenker and Walker, 2016), the original high-frequency flight-by-flight dataset used in this study has a considerable number of missing observations on actual departure times. To deal with this issue, we use the flight status information provided by the raw database to fill in some missing values. For flights that have missing values on actual departure times but have flight status recorded as "regular, no delay", we assume no delay occurred and recode the actual departure time as the scheduled departure time.

Flight departure delays are typically measured by the difference between a flight's actual and scheduled departure times (Brueckner, 2005; Deshpande and Arıkan, 2012; Forbes, 2008; Mazzeo, 2003). However, this approach may lead to an overestimation of flight delays because it does not rule out delays caused by a flight's previous assignments in other airports. Here, we use two methods to measure flight delays. First, we calculate a flight's *gross* departure delay based on its actual and scheduled departure times. Second, for flights with previous assignments, we deduct these flights' arrival delays from the *gross* delays to calculate *net* flight delays, which are solely related to conditions in the current departure airports. We compute a flight's arrival delay based on its actual and scheduled arrival times.

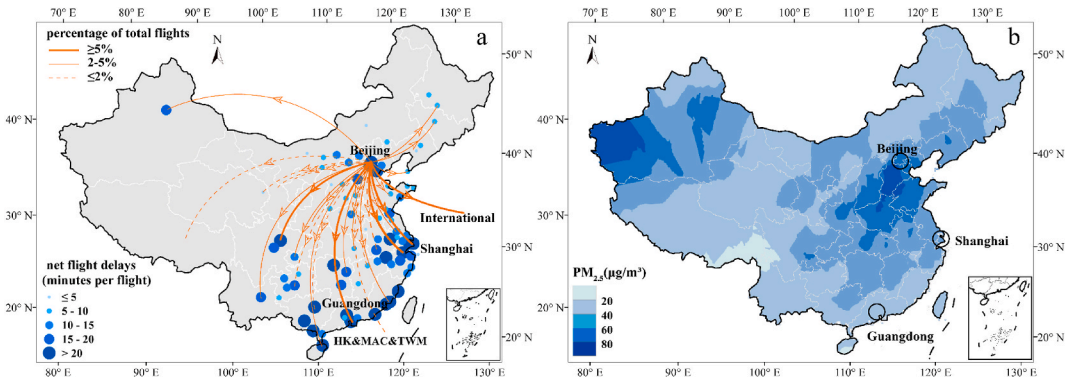



**Fig. 1.** Airline travel and $PM_{2.5}$ concentration in China during 2015–2017. Note: **a:** The percentage of flight takeoffs from North China and net departure delays for airports covered by our sample. **b:** The spatial distribution of $PM_{2.5}$ concentrations.

We focus on estimating the effect of pollution on the daily average flight delay at the airport level, which is equal to the sum of the delays of all flights throughout each day, divided by the daily total number of flights. To ensure that our estimates are not biased due to outliers, we drop observations from the raw data set if the net delay per flight is either above the 99th percentile or below the 1st percentile. We also drop all flights operated by cargo companies that carry packages only.

### 2.2. Air pollution data

We obtain air pollution data from the China National Environmental Monitoring Center, which provides hourly readings of ground-level concentrations of six air pollutants: sulfur dioxide ($SO_2$), carbon monoxide (CO), ozone ($O_3$), nitrogen dioxide ($NO_2$), and particulate matter ($PM_{10}$ and $PM_{2.5}$) for approximately 1600 ground monitoring stations, beginning in 2013. These monitoring stations are widely distributed across China, with the number of monitoring stations gradually decreasing from east to west (see Figure A1 in Greenstone et al., 2022). To match up with our flight data, we convert the hourly pollution data into the daily averages by taking the duration-weighted average of all hourly readings. We then use an inverse distance weighting (IDW) method to impute pollution concentrations for our sample airports. Specifically, we choose a radius of 50 km surrounding the centroid of each airport and compute the distance-weighted averages of pollution concentrations recorded by all monitoring stations within the circle. Our results are robust to alternative radii.

### 2.3. Weather and satellite-based thermal inversion data

The weather data are obtained from the China Meteorological Data Service Center (http://data.cma.cn/), which reports daily maximum, minimum and average temperatures (in degrees Celsius, °C), total daily precipitation (in millimeters, mm), wind speed (in meters per second, m/s), wind direction (in degrees), water vapor pressure (in hPa), air pressure (in hPa), relative humidity (in %) and visibility (in meters, m) for approximately 820 weather stations in mainland China. Wind direction is defined as the direction from which the wind is blowing. We construct airport-level weather information by using the IDW method in the same fashion as the pollution variables are generated.

We obtain thermal inversion data from the U.S. National Oceanic and Atmospheric Administration, which provides the Modern-Era Retrospective analysis for Research and Applications (MERRA) satellite data containing air temperatures every 3 h at 42 vertical levels from the surface to 36 thousand meters.[4] The MERRA data are provided with a spatial resolution of $0.5^\circ \times 0.625^\circ$ (approximately 45 km × 55 km). We extract these grid-level inversion data for China between 2015 and 2017. An airport is assigned the readings from the MERRA coordinate with the closest proximity to that airport.

Following Arceo, Hanna, and Oliva (2016), we define a thermal inversion as the air temperature of the second layer (925 hPa, about 320 m) being higher than that of the first layer (1000 hPa, about 110 m). We use the strength of thermal inversions, measured as the daily aggregated temperature difference between the second and first layer, as our primary instrument. We also measure thermal inversions using (*i*) a dummy variable that takes the value of 1 if thermal inversions occur on a day and 0 otherwise, and (*ii*) the total number of hours with thermal inversions on a day to measure the frequency of thermal inversions.

Table 1 reports descriptive statistics for key variables. The final sample consists of 74,122 airport-day observations. The daily average gross delay per flight is 35.7 min, which is significantly larger than the daily average net delay per flight of 13.6 min. The mean daily concentration of $PM_{2.5}$ is 45.3 μg/m$^3$. On average, inversion episodes occur during 7.5% of a day (= 1.8 h/24 h). Other pollution and temperature variables also exhibit considerable variations during the sample period.

## 3. Empirical methodology

Our objective is to estimate the causal effect of $PM_{2.5}$ on flight delays, net of any potentially confounding factors. There exist several identification challenges. The primary challenge stems from the fact that air pollution is not randomly assigned and is subject to measurement errors. Regulators also may have strategically chosen the locations of monitoring stations. In our case, both flight departure delays and pollution may be influenced by unobserved confounding factors. We tackle these econometric challenges by taking an instrumental variable approach, with thermal inversions and wind directions as two separate instruments. Following prior studies in the literature explaining the reasons for flight delays (Mayer and Sinai, 2003; Mazzeo, 2003), we employ a linear specification to estimate the relationship between pollution and flight delays.

### 3.1. Panel fixed effects model

We first model the relationship between $PM_{2.5}$ and flight delays using the following fixed effect panel regression approach:

$$Y_{i,t} = \beta P_{i,t} + X_{i,t}^{'}\gamma + \theta_t + c_i + \varepsilon_{i,t} \quad (1)$$

where the dependent variable, $Y_{i,t}$, measures the average flight departure delay (minutes per flight) at departing airport *i* on day *t*. $P_{i,t}$ is the average daily $PM_{2.5}$ concentration. The parameter of interest is $\beta$, which captures the effect of $PM_{2.5}$ on flight departure delays. We

[4] https://disc.gsfc.nasa.gov/datasets/M2I3NVAER_V5.12.4/summary. We use the product M2I3NVAER version 5.12.4.

**Table 1**
Descriptive statistics.

| Variables | Mean | Standard deviation |
|---|---|---|
| Daily Average Flight Delays (minutes per flight) | | |
| Net flight delays | 13.6 | 23.6 |
| - *Domestic flights* | 13.6 | 23.6 |
| - *International flights* | 26.5 | 30.2 |
| Gross flight delays | 35.7 | 39.6 |
| **Daily Average Air Pollution** | | |
| Particulate matter ($PM_{2.5}$) ($\mu g/m^3$) | 45.3 | 37.4 |
| Sulfur dioxide ($SO_2$) ($\mu g/m^3$) | 20.9 | 23.1 |
| Ozone ($O_3$) ($\mu g/m^3$) | 61.0 | 30.6 |
| Carbon monoxide (CO) ($mg/m^3$) | 1.0 | 0.6 |
| Nitrogen dioxide ($NO_2$) ($\mu g/m^3$) | 34.2 | 19.3 |
| **Daily Weather** | | |
| Average temperature (°C) | 16.5 | 10.4 |
| Total precipitation (mm) | 3.1 | 10.5 |
| Average wind speed (m/s) | 5.7 | 2.8 |
| Water vapor pressure (hPa) | 15.3 | 9.4 |
| Atmospheric pressure (hPa) | 974.2 | 56.9 |
| Relative humidity (%) | 68.6 | 18.3 |
| Visibility (m) | 18240.5 | 12275.7 |
| **Thermal Inversions** | | |
| Dummy: whether thermal inversions occur | 0.3 | 0.5 |
| Frequency: number of hours with thermal inversions in each day (hours) | 1.8 | 3.6 |
| Strength: aggregated temperature difference between the first and the second layers (K) | 35.0 | 15.4 |

Note: Sample size: 74,122, including 89 airports over the period April 1, 2015–September 30, 2017.

hypothesize that increased $PM_{2.5}$ concentrations are likely to increase ground congestion at airports, possibly by reducing atmospheric visibility and/or airport operational efficiency. The latter occurs because transitory $PM_{2.5}$ exposure can negatively affect the cognitive performance of airport ground and cabin crews. Hence, $\beta$ is expected to be positive.

In addition to $PM_{2.5}$, short-term exposure to other air pollutants, such as CO, $SO_2$, $O_3$ and $NO_2$, can adversely impact human health and reduce productivity (Gao et al., 2021; Zhao et al., 2018). $SO_2$ and $NO_2$, which are precursors to $PM_{2.5}$, as well as $O_3$ play central roles in smog formation. Throughout most of this paper, we focus on assessing the impact of $PM_{2.5}$ on flight delays for two reasons. One, as the primary air pollutant in China, ambient $PM_{2.5}$ levels routinely exceed the World Health Organization (WHO) standard level for good health. $PM_{2.5}$ has been identified as the principle pollutant to tackle through China's clean air policies (Chen and Ye 2019). Two, $PM_{2.5}$ has been used almost exclusively as a measure of air pollution in China (He et al., 2019; Zhang et al., 2018, 2019). However, we also control for other air pollutants and examine whether this alters our estimates of the $PM_{2.5}$ impact on flight delays (see Table 6 in Section 4.4) .

Flexibly controlling for weather conditions is crucial when investigating the impact of air pollution on flight delays, because weather is a key factor influencing air traffic delays in the aviation system (Borsky and Unterberger, 2019). This becomes especially prominent when using thermal inversions as an instrument, because inversions are related to changes in temperature both at ground level and higher up in the troposphere. To this end, we generate indicator variables for daily average temperatures, which fall into one of nine temperature bins, with each bin 5 °C wide and ranging from below 0 °C to above 35 °C. Weather variables, denoted by $X'_{i,t}$ in equation (1), also include indicators for deciles of daily precipitation, as well as daily average wind speed, water vapor pressure, air pressure, and relative humidity. These variables have been identified as the key weather factors that can affect flight delays (Borsky and Unterberger, 2019). Moreover, the impacts of weather on flight departure delays may be nonlinear and depend on the interactions of some weather variables. Thus, we consider several alternative specifications. We first include all possible interaction terms of temperature (5 °C bins) and precipitation (decile bins) indicators, in addition to the other weather covariates included in the baseline specification. Second, we generate indicators for deciles of wind speed and relative humidity, respectively, and find that our estimates are robust to more flexible weather controls that incorporate all possible interaction terms of these temperature, precipitation and wind (or humidity) variables (see Section 4.3).

$\theta_t$ is a vector of time fixed effects that are used to capture temporal shocks to flight delays, including year, month and day-of-the-week fixed effects. $c_i$ denotes airport fixed effects, capturing time-persistent unobserved airport attributes that affect flight delays, such as geographical locations, operation strategies, and systems. Since most Chinese cities have only one airport, the airport fixed effects can also absorb city-specific time-invariant factors that might affect flight delays. $\varepsilon_{i,t}$ are the error terms capturing time-varying, airport-specific unobservables that may affect flight delays. The error terms may be spatially and temporally correlated. To account for this, the regression is estimated by clustering the standard errors in two dimensions: within airports and within air traffic control region-days. The former accounts for serial correlation within each airport, while the latter accounts for spatial correlation across

airports within each air traffic control region.[5] We also control for the heteroskedasticity of the error terms.

As noted above, the OLS estimator of $\beta$ is prone to bias for three reasons: (*i*) $PM_{2.5}$ is not randomly assigned across regions; (*ii*) pollution data may be subject to measurement errors and manipulation (Ghanem and Zhang, 2014); and (*iii*) airports are a main source of air pollution. For instance, airport runway congestion, caused by network delays or air traffic controls, may significantly increase local air pollution (Schlenker and Walker, 2016). Moreover, flight delays and the pollution level at a certain airport may be related to unobserved confounding factors. For instance, air traffic volume can affect flight departure delays and air pollution at the same time. Therefore, a naïve simple linear regression estimate might bias the true effect of pollution on flight delays.

### 3.2. Instrumental variable models

We address these econometric challenges by using two separate instruments that rely on changes in atmospheric thermal inversions and wind direction as exogenous shocks to local pollution levels. We first use thermal inversions as the instrument. Air temperature typically decreases with an increase in altitude. A thermal inversion is a deviation from this atmospheric norm. During an inversion, a layer of warmer air is held above a layer of cool air, trapping pollutants on the ground and resulting in an increase in pollution concentrations. Thermal inversions have been widely used as an instrument for pollution to explore the effects of pollution on health (Arceo et al., 2016; Jans et al., 2018), productivity (Fu et al., 2021; He et al., 2019), road safety (Sager, 2019) and crime (Bondy et al., 2020). We denote $INVERSION_{i,t}$ as the inversion instrument and specify our first stage as follows:

$$P_{i,t} = \pi INVERSION_{i,t} + X'_{i,t}\varphi + \theta_t + c_i + \sigma_{i,t} \tag{2}$$

There are at least three ways of constructing the inversion instrument, including inversion strength, a binary indicator, and inversion frequency. As shown in the robustness check section, the regression results are highly consistent across alternative measures of inversions (Appendix Table S4). Other control variables and the fixed effects in equation (2) are defined as in equation (1).

We also use wind directions as an alternative instrument for pollution. More specifically, we estimate the following first stage model:

$$P_{i,t} = \sum_{g\in G}\sum_{a=0}^{2} \pi_a^g \mathbf{1}[G_i = g] \times WindDirection_{i,t}^{90a,90a+90} + X'_{i,t}\varphi + \theta_t + c_i + \sigma_{i,t} \tag{3}$$

The variable $\mathbf{1}[G_i = g]$ is an indicator for airport $i$ being assigned to group $g$ from the set of air quality monitoring group $G$. Following the recent work of Deryugina et al. (2019), Bondy et al. (2020), and Carneiro et al. (2021), we use the $k$-means cluster algorithm to generate 100 groups for all the pollution monitors in China based on their coordinates. The variable $WindDirection_{i,t}^{90a,90a+90}$ is another indicator, which is equal to 1 if the daily average wind direction at airport $i$ on day $t$ falls in the 90-degree interval [90a, 90a +90) and 0 otherwise. We choose the interval [270, 360) as the reference category. The interaction term $\mathbf{1}[G_i = g] \times WindDirection_{i,t}^{90a,90a+90}$ thus contains our excluded instruments. Our results are robust to variations in the numbers of spatial groups and wind direction bins (see Table 4 in Section 4.3)(. The coefficient $\pi_a^g$ captures the influence of wind direction on pollution and is allowed to vary across regions. We do so because airports covered by our sample are widely distributed and wind may carry pollution from different sources in upwind regions, resulting in the same wind direction having differential effects on pollution at different airports. Other control variables and the fixed effects are constructed as in equation (1).

IV estimates can be interpreted as a local average treatment effect (LATE) (Imbens and Angrist, 1994). But this interpretation requires that the chosen instruments meet three identifying assumptions, namely instrument relevance, monotonicity and exclusion restriction (or independence). In our setting, the instrument relevance requires that both inversions and wind direction affect pollution concentrations. The first-stage Kleibergen-Paap F-statistics reported in Table 2 are larger than 25, indicating that inversions and wind direction are indeed relevant instruments.

When using wind direction as the instrument, the monotonicity assumption requires that an airport's $PM_{2.5}$ level always increases (decreases) when the wind blows from a high (low) pollution direction. Monotonicity may be violated if some airports in a monitor group exhibit different responses to wind than do the others within the same group, or if the relationship between $PM_{2.5}$ and wind direction changes within a 90-degree interval. To probe these possibilities, following Deryugina et al. (2019), we estimate alternative specifications by considering a wide range of variations in the numbers of spatial monitor groups and/or wind direction bins (see results in Table 4 in Section 4.3). For the inversion instrument, the monotonicity assumption is less of a concern, because a thermal inversion is a weather phenomenon, and it is unlikely that the pollution-inversion relationship holds in some regions (or times of year) but is reversed in other regions (or times of year).

Finally, the exclusion restriction (independence) assumption requires that inversions and wind direction are not directly related to the changes in flight departure delays except through their influences on pollution levels. This assumption may be violated because thermal inversions may be accompanied by other adverse weather conditions, such as fog or freezing rain, which directly affect flight planning. Similarly, wind direction can also directly affect flight departure delays, although the impact is small (Borsky and Unterberger, 2019). Apart from flexibly controlling for ground-level weather conditions, following Sager (2019), we conduct several

[5] The CAAC divides mainland China into seven air traffic control regions: Northwest, Northeast, Central South, North, East, Southeast, and Xinjiang regions. The subdivision of CAAC in each of these regions is responsible for flight coordination within its own territory.

robustness checks by (*i*) incorporating all possible interactions of temperature and precipitation variables, (*ii*) incorporating all possible interactions of temperature, precipitation, and wind speed (or humidity) variables, (*iii*) including an additional control for visibility, (*iv*) excluding "red-alert" days from the full sample,[6] (*v*) excluding foggy days from the full sample, and (*vi*) using a sub-sample with flight conditions recorded as "good" for flight departures, to ensure that the independence assumption is not endangered. Reassuringly, the results based on a range of robustness checks provide supportive evidence of the robustness of our estimates to the presence of unobserved confounding factors (see Table 5 in Section 4.3).

## 4. Results

### 4.1. Baseline results

Table 2 reports on the relationship between daily $PM_{2.5}$ pollution and flight delays. Columns 1–2 report the OLS estimates. In column 1, we include the fixed effects discussed in equation (1), while column 2 adds weather covariates. The coefficient estimates in both columns suggest a small and positive correlation between $PM_{2.5}$ and flight delays. Each 1 $\mu g/m^3$ increase in daily average $PM_{2.5}$ concentration is associated with an increase in net delay per flight by 0.004 min and an increase in gross delay per flight by 0.037 min. The estimates on net flight delays are not statistically significant, but the estimates on gross flight delays are statistically significant at the 1% level. However, the OLS estimates are subject to a range of biases discussed above, preventing a causal interpretation.

Columns 3 and 4 of Table 2 present the IV estimates of the causal effect of daily $PM_{2.5}$ pollution on flight delays, using strength of thermal inversions as the instrument. Appendix Table S1 shows that each additional 1 °C increase in the difference in temperature between the second layer (the 925 hPa pressure level) and the surface level is associated with an increase in $PM_{2.5}$ concentrations by 0.31–0.36 $\mu g/m^3$. The first stage Kleibergen-Paap F-statistics are greater than 25, suggesting that inversions are indeed a relevant instrument for daily $PM_{2.5}$ concentrations. Column 3 reports the IV estimates without weather controls. The estimated impacts of $PM_{2.5}$ on flight delays are substantially (approximately 10–50 times) larger than the corresponding estimates in column 1, suggesting that the OLS estimates suffer from significant bias. The IV estimates in column 3 imply that each 1 $\mu g/m^3$ increase in daily average $PM_{2.5}$ exposure is associated with an increase of 0.217 min of net delay per flight and an increase of 0.396 min of gross delay per flight. These coefficients are statistically significant at the 1% level. Column 4 reports the IV estimates by adding weather controls. We find that the signs and levels of statistical significance of the coefficient estimates are broadly consistent with those in column 3. However, the magnitudes of the estimated coefficients become 50–60% smaller: each 1 $\mu g/m^3$ increase in daily $PM_{2.5}$ is associated with an increase in net delay per flight by 0.110 min and an increase in gross delay per flight by 0.159 min. This finding indicates the importance of controlling for weather variables, since weather plays a key role in influencing air traffic delays in the aviation system (Borsky and Unterberger, 2019).

**Table 2**
Estimates of the causal effects of $PM_{2.5}$ on flight delays.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | OLS | OLS | 2SLS (Inversions) | 2SLS (Inversions) | 2SLS (Wind direction) | 2SLS (Wind direction) |
| Panel A: Net delay per flight | | | | | | |
| $PM_{2.5}$ | 0.00425 | 0.00431 | 0.217*** | 0.110*** | 0.114*** | 0.119*** |
| | (0.00554) | (0.00467) | (0.0509) | (0.0398) | (0.0385) | (0.0352) |
| Panel B: Gross delay per flight | | | | | | |
| $PM_{2.5}$ | 0.0368*** | 0.0366*** | 0.396*** | 0.159** | 0.180** | 0.170** |
| | (0.0101) | (0.00904) | (0.0851) | (0.0694) | (0.0735) | (0.0671) |
| Observations | 74,122 | 74,122 | 74,122 | 74,122 | 74,122 | 74,122 |
| F-test (KP statistics) | – | – | 26.78 | 39.67 | 344.8 | 64.91 |
| Weather controls | – | YES | – | YES | – | YES |

Note: This table reports the estimated impacts of daily average $PM_{2.5}$ level on daily average net delay per flight and gross delay per flight, respectively. Columns 1–2 report the ordinary least squares (OLS) estimates. Columns 3–4 report the estimated impacts of $PM_{2.5}$ using the strength of thermal inversions (aggregated temperature difference between the second and the first layers) as the instrument for $PM_{2.5}$. Columns 5–6 report the estimates using wind direction as the instrument for $PM_{2.5}$. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.

[6] In 2013, the Chinese governments initiated a red-alert system to keep people indoors on days with poor air quality and low visibility. Specifically, a red alert is issued when air quality index reaches over 300 and is projected to last for at least three consecutive days. Beijing issued the first red alert on December 7, 2015. Since then, the majority of red alerts have been issued in prefecture-level cities in smog-plagued regions in Northern China. An issued red alert triggers a series of compulsory measures, including limiting car use, shutting down outdoor construction operations, advising schools to close, and even closing some industrial firms. These compulsory measures may also affect flight planning and operations.

**Table 3**
Robustness of IV estimates using wind direction as an instrument.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | Adding province × year FE | Adding year × month FE | Clustering SD within airports and dates | IDW within 25 km | IDW within 100 km | Removing flights between 12 a.m. and 6 a.m. | Controlling $PM_{2.5}$ at landing airports | PPML | LIML |
| $PM_{2.5}$ (departure airports) | 0.114*** | 0.120*** | 0.119*** | 0.120*** | 0.0867*** | 0.119*** | 0.0806* | 0.00723*** | 0.149*** |
| | (0.0359) | (0.0372) | (0.0348) | (0.0389) | (0.0296) | (0.0352) | (0.0440) | (0.00245) | (0.0372) |
| $PM_{2.5}$ (landing airports) | | | | | | | 0.00325 | | |
| | | | | | | | (0.00466) | | |
| Observations | 74,122 | 74,122 | 74,122 | 49,160 | 86,608 | 74,122 | 436,740 | 74,122 | 74,122 |
| F-test (KP statistics) | 441.7 | 74.40 | 32.12 | 881.1 | 113.1 | 64.91 | 582.9 | – | 64.91 |

Note: This table reports estimated impacts of daily average $PM_{2.5}$ level on daily average net delay per flight with alternative specifications, data and estimation strategies, using daily wind direction as the instrument. Columns 1-2 additionally include province × year FE and year × month FE, respectively. Column 3 clusters the standard errors at the airport and date levels to further address concerns on potential network effects of flight delays across regions. Columns 4–5 use a radius of 25 km and 100 km, respectively, to generate distance-weighted pollution and weather data for sample airports. Column 6 excludes overnight flights between 12 a.m. and 6 a.m. in the sample. Column 7 reports estimates from the bilateral-route specification which includes the daily average $PM_{2.5}$ level and weather controls at both the departing and landing airports. Column 8 uses a Poisson pseudo maximum likelihood (PPML) specification and bootstraps standard errors with 1000 replications, while the endogeneity of $PM_{2.5}$ is addressed by using control function method. Column 9 uses the limited information maximum likelihood (LIML) estimator instead of the 2SLS estimator. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level except in columns 3 and 8. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Columns 5 and 6 of Table 2 report estimates from another IV strategy, where we use wind direction as an instrument for daily $PM_{2.5}$ concentrations. The large first-stage Kleibergen-Paap F-statistics indicate that wind direction is also a strong predictor of local $PM_{2.5}$ concentrations. The second-stage estimates in the two columns are also economically and statistically significant and have similar magnitudes. In our preferred specification incorporating weather covariates (in column 6), we find that, holding all else equal, each 1 $\mu g/m^3$ increase in daily $PM_{2.5}$ concentration increases net delay per flight by 0.119 min and gross delay per flight by 0.170 min. These coefficient estimates are also remarkably close to those obtained when using inversions as the instrument (column 4).

The similarity in the estimated coefficients of the $PM_{2.5}$ variables when using the two IV strategies strengthens our confidence that our results yield credible evidence on an arguably causal link between $PM_{2.5}$ and flight delays. We next translate the point estimates displayed in column 6 into changes in flight delays per one-standard-deviation increase in daily $PM_{2.5}$. We find that, for a one-standard-deviation increase in daily $PM_{2.5}$, net delay per flight increases by 4.5 min (= 37.4*0.119) and gross delay per flight increases by 6.4 min (= 37.4*0.170).

Coefficients on weather variables are reported in Appendix Table S2. We find that precipitation exerts a strong impact on flight delays and the impact increases with the level of precipitation. Specifically, when the daily precipitation level is above the 90th percentile in our sample, net delay per flight is expected to increase by an additional 9.4–9.5 min and gross delay per flight is expected to increase by 14.3–14.5 min, holding all else equal. This impact is considerably larger than the effects of other weather controls on flight departure delays.

### 4.2. Robustness checks

We conduct a number of robustness checks and placebo exercises to test the sensitivity of our baseline estimates. Because *gross* flight delays do not rule out delays caused by a flight's previous assignments in other airports, this measure overestimates true delays that occurred at the departing airports. Hence, we use net delay per flight as the dependent variable in the remainder of the paper, while choosing the specification that incorporates weather covariates.

Table 3 shows that our result remains robust to different types of fixed effects, clustering strategy, and how airport-level weather and pollution variables are generated, with $PM_{2.5}$ instrumented by wind direction. Columns 1–2 of Table 3 report the coefficient estimates by additionally adding province × year FE and year × month FE, respectively. These are intended to control for unobserved confounders varying over time in a region, such as air traffic control, seasonality in air travel demand, and so on. We find that, after accounting for these FE, the coefficient estimates are nearly identical to our baseline estimate, providing strong evidence that our results cannot be explained by regional or seasonal patterns. Column 3 estimates the regression by clustering the standard errors within airports and within dates to address the concern that there may exist substantial network effects of flight delays across regions. Reassuringly, the estimated $PM_{2.5}$ impact on net flight delay remains unchanged.

In the baseline regression, we select a radius of 50 km to generate distance-weighted pollution and weather data for each airport in our sample. Columns 4–5 report the coefficient estimates when changing the radius from 50 km to 25 km and 100 km, respectively. The coefficient estimates in columns 4–5 are again in agreement with our baseline estimates, suggesting that our baseline estimates can be interpreted as the causal effect of each one-unit change in daily $PM_{2.5}$ on flight delay.

Some severely delayed overnight flights might have been canceled, and this information is not recorded in the dataset. To ensure that our estimates are not driven by canceled overnight flights, we exclude overnight flights between 12 am and 6 am from the analysis. Column 6 in Table 3 shows that this yields an estimate similar to the baseline estimate: a one-standard-deviation increase in daily $PM_{2.5}$ level leads to an increase of 4.5 min (=37.4*0.119) in net delay per flight.

Intuitively, pollution and weather conditions at destination airports could also affect flight planning at departing airports and hence lead to delay or even cancellations of flights. Therefore, our baseline estimates, which focus only on pollution and weather conditions at the departure airports, might have yielded biased estimates of the true effect of pollution on flight delays. Column 7 of Table 3 extends our baseline specification by incorporating $PM_{2.5}$ and weather at the landing airports as covariates. This allows us to separately identify the effect of $PM_{2.5}$ at both the departure and destination airports.[7] Specifically, we modify equation (1) to the following specification:

$$Y_{i,j,t} = \beta_1 P_{i,t} + X'_{i,t}\gamma_1 + \beta_2 P_{j,t} + X'_{j,t}\gamma_2 + \theta_t + c_{i,j} + \varepsilon_{i,t}$$

where the dependent variable, $Y_{i,j,t}$, measures the daily average departure delay for flights departing from airport $i$ and landing at airport $j$. $P_{j,t}$ and $X'_{j,t}$ are pollution and weather conditions at airport $j$ on day $t$. $c_{i,j}$ denotes bilateral-route fixed effects, which can control for time-invariant unobserved attributes between a pair of airports, such as geography, which could affect flight delays. Other variables are defined as in equation (1). As shown in Column 7 of Table 3, this yields a smaller estimate relative to the baseline result. Holding all else equal, each 1 $\mu g/m^3$ increase in daily $PM_{2.5}$ concentration at the departure airports is associated with a net delay per flight of 0.0806 min, which can be translated into a net flight delay of 3.0 min for each one-standard-deviation increase in daily $PM_{2.5}$ level. The estimated coefficient for the $PM_{2.5}$ variable at the destination airports is positive and statistically insignificant. The magnitude of this estimate is also small, suggesting that $PM_{2.5}$ pollution at the departure airports plays a more important role in affecting flight departure delays, relative to the pollution concentrations at the destination airports.

[7] We are indebted to an anonymous referee for suggesting route-level analysis as an alternative empirical design to separate the pollution effects at the departure and destination airports.

Further, the relationship between pollution and flight delays may be nonlinear. To probe this, we employ the Poisson pseudo maximum likelihood (PPML) estimation as an alternative specification to examine the possible nonlinear relationship between pollution and flight delays.[8] We bootstrap standard errors with 1000 replications and address the endogeneity of $PM_{2.5}$ by using a control function method. Column 8 of Table 3 reports that each additional 1 $\mu g/m^3$ increase in daily $PM_{2.5}$ concentration is associated with a 0.7% increase in net flight delays, and this estimate is statistically significant at the 1% level. Our baseline estimate shows that each 1 $\mu g/m^3$ increase in daily $PM_{2.5}$ is associated with an increase in net delay per flight by 0.119 min, which can be translated into a percentage increase of 0.88%, with an average net delay per flight of 13.6 min in our sample. Hence, the PPML estimate is again in agreement with our baseline result.

Moreover, we conduct three additional analyses to ensure the robustness of our baseline estimate. First, we estimate equation (1) using the limited information maximum likelihood (LIML) estimator, which is approximately median unbiased with weak instruments (Staiger and Stock, 1997). Column 9 of Table 3 reports the LIML estimate is similar to those obtained from the 2SLS specification, suggesting that our estimate does not suffer from weak instrument bias. Second, we perform a set of placebo tests where we randomly generate wind directions for each airport 500 times and use those, rather than the actual wind directions, as instruments in our first stage. Appendix Fig. S2 plots the distribution of estimated placebo treatment effects from the 500 random assignments. The dashed line in this figure denotes the estimated effect of $PM_{2.5}$ from the baseline analysis. The $p$ value of the placebo test is 0.02, which rejects the null hypothesis of no effect of $PM_{2.5}$ and provides further support for our identification strategy. Third, we conduct a placebo exercise in which we estimate the relationship between pollution and flight delays on irrelevant days. Specifically, we replace the $PM_{2.5}$ variable in equations (1) and (3) with one- and three-day leads of pollution level.[9] Appendix Table S3 reports that there are no significant relationships between the placebo pollution readings and flight delays.

Columns 1–2 of Appendix Table S4 show that the estimated pollution impact on flight delays remains broadly consistent to our baseline result when using the frequency of thermal inversions or a dummy indicator of thermal inversions as the instrument for the concentration of $PM_{2.5}$. In columns 3–10 of Appendix Table S4, we replicate the robustness checks described above, using the strength of inversions as an instrument for $PM_{2.5}$. We find that these estimates are again similar to the baseline estimate (column 4 of Table 2).

### 4.3. Extensions

***Testing the validity of the monotonicity assumption:*** As noted above, the monotonicity assumption is less of a concern for the inversion instrument. To investigate the validity of the monotonicity assumption when using wind direction as the instrument for $PM_{2.5}$ we estimate alternative specifications by permitting instruments to vary with the size of wind direction bins and the number of monitor groups.[10] For ease of comparison, column 1 of Table 4 reports our baseline estimate. Columns 2 and 3 decrease the size of the wind direction bins from 90 to 45 and 60°, respectively. Column 4 increases the size of the wind direction bins to 120°. Column 5 decreases the number of monitor groups from 100 to 50. Table 4 demonstrates that our results are generally robust to these variations. The stability of the coefficient estimates lends support to the monotonicity assumption for the wind direction instrument. If a violation of this assumption exists, it is unlikely to affect our ability to interpret our results as LATE.

**Table 4**
Robustness of IV estimates to instrument choices.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $PM_{2.5}$ | 0.119*** | 0.108*** | 0.0987*** | 0.132*** | 0.102** |
| | (0.0352) | (0.0271) | (0.0218) | (0.0367) | (0.0446) |
| Observations | 74,122 | 74,122 | 74,122 | 74,122 | 74,122 |
| F-test (KP statistics) | 64.91 | 133.4 | 245.6 | 50.50 | 87.18 |
| Num. Of pollution monitor groups | 100 | 100 | 100 | 100 | 50 |
| Size of wind angle bins (degrees) | 90 | 45 | 60 | 120 | 90 |

Note: This table reports estimated impacts of daily average $PM_{2.5}$ level on daily average net delay per flight, using wind direction as the instrument for $PM_{2.5}$. Column 1 reports the baseline estimate that separates pollution monitoring stations into 100 groups and wind angles into 90-degree intervals. Columns 2–5 report specifications using alternative numbers of monitor groups and wind angle bins. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.

[8] There are two major reasons why we adopt the PPML rather than other non-linear estimation methods. First, due to the presence of many zero values in our main dependent variables, traditional log-transformation and log-linear OLS estimations will exclude a fairly large portion of our observations (over 20%). PPML can make use of zeros and provide consistent and robust estimates in this case. Second, PPML can be easily implemented in our case because it does not require a distributional assumption for the dependent variable (e.g., count data, etc.). More technical details can be found in Correia et al. (2020).

[9] Here, we do not replace the $PM_{2.5}$ variable with lagged pollution levels, because $PM_{2.5}$ exhibits significant temporal correlations (Chen and Ye, 2019).

[10] Deryugina et al. (2019) conduct a similar robustness exercise to examine the impacts of air pollution on mortality in the U.S.

***Testing the validity of the exclusion restriction (independence) assumption:*** The core of the exclusion restriction assumption requires that inversions and wind direction affect flight departure delays only through their influences on pollution concentrations. As discussed above, both instruments may be related to other weather events that also affect flight delays. We conduct several analyses to demonstrate the validity of the exclusion restriction assumption in our setting. Apart from flexibly controlling for ground-level weather conditions, following Sager (2019) we conduct several robustness checks by (*i*) incorporating all possible interactions of temperature and precipitation variables, *(ii)* incorporating all possible interactions of temperature, precipitation, and wind speed (or humidity) variables, (*iii*) including an additional control for visibility, (*iv*) excluding days with red alerts from the full sample, (*v*) excluding days with fog or freezing rain from the full sample, and (*vi*) using a sub-sample with flight conditions recorded as "good" for flight departures, to ensure that the independence assumption is not endangered. Table 5 reports the results, with $PM_{2.5}$ instrumented by wind direction.

Column 1 of Table 5 includes all possible interaction terms of temperature (5 °C bins) and precipitation (decile bins) indicators, in addition to the weather covariates included in the baseline specification. Column 2 adds all possible interaction terms of temperature (5 °C bins), precipitation (decile bins) and deciles of wind speed. Column 3 adds all possible interaction terms of temperature (5 °C bins), precipitation (decile bins) and deciles of humidity variables.[11] We find that the estimated coefficients become slightly smaller after controlling for a full set of weather interaction terms but are still consistent with the baseline estimate.

As noted above, we have hypothesized that air pollution may cause flight delays or cancellations by reducing atmospheric visibility and/or impairing cognitive performance. We now provide suggestive evidence about these possible mechanisms. Adding visibility as a control helps to disentangle the channels through which $PM_{2.5}$ affects flight delays. Column 4 of Table 5 shows that adding the daily average visibility as a control does not substantially alter our results. The coefficient estimated in column 4 is again statistically significant at the 1% level. To examine whether the results are influenced by government-mandated actions, we exclude the days with red alerts from the full sample.[12] Due to the concern about the network effects of flight delays across regions, we exclude the days from the full sample if red alerts were issued in any of the sample cities during the study period. Column 5 of Table 5 reports the coefficient estimate from analyzing this sub-sample, which is again similar to our baseline estimate. Together, these findings in columns 4 and 5 rule out poor visibility as an alternative explanation and suggest that impaired cognitive performance may be the primary factor causing flight departure delays.

In addition, it is noted in the literature that the formation of fog might be linked to thermal inversion events and wind direction/speed, especially when wind blows from a moisture source (Gultepe et al., 2007). To ensure that our estimates are not driven by

**Table 5**
Robustness of IV estimates to weather controls and weather conditions.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Flexible weather controls for temperature and precipitation | Flexible weather controls for temperature, precipitation, and wind speed | Flexible weather controls for temperature, precipitation, and relative humidity | Adding visibility as an additional control | Excluding "red-alert" days | Excluding foggy days | Flight conditions recorded "good" for departures |
| $PM_{2.5}$ | 0.111*** | 0.106*** | 0.0972*** | 0.183*** | 0.106*** | 0.0784*** | 0.0970*** |
| | (0.0354) | (0.0357) | (0.0349) | (0.0261) | (0.0391) | (0.0295) | (0.0286) |
| Observations | 74,122 | 74,122 | 74,122 | 74,122 | 72,061 | 51,215 | 72,575 |
| F-test (KP statistics) | 54.30 | 61.19 | 27.41 | 108.8 | 64.908 | 45.62 | 81.73 |

Note: This table reports estimated impacts of daily average $PM_{2.5}$ level on daily average net delay per flight with alternative weather controls and sub-samples, using wind direction as the instrument for $PM_{2.5}$. Columns 1–3 show the results equivalent to our baseline specification, but with more flexible weather controls. Column 1 includes all possible interactions for temperature (5 °C bins) and decile bins for precipitation. Column 2 includes all possible interactions for temperature (5 °C bins), precipitation (decile bins), and wind speed (decile bins). Column 3 includes all possible interactions for temperature for temperature (5 °C bins), precipitation (decile bins), and relative humidity (decile bins). Column 4 includes daily average airport-level atmospheric visibility as an additional control. Column 5 reports the estimate from a sub-sample that excludes days with red alerts. Column 6 reports the estimate from a sub-sample of days with daily total precipitation below 1 mm and relative humidity below 85%. Column 7 reports the estimate from a sub-sample when flight conditions were recorded "good" for flight departures, which excludes days with heavy rainfall, high relative humidity, low visibility and flight controls due to military activities. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.

[11] Ideally, we will have up to 899 (=9*10*10–1) weather indicators containing interaction terms of temperature, precipitation, and wind speed (or humidity) in columns 2 and 3. However, not all possible interaction terms are realized in practice, and the actual numbers of weather controls in the two columns are about 870 and 770, respectively.

[12] We are indebted to an anonymous referee for suggesting this analysis. We hand-collected red-alert data from various local Environmental Protection Bureau portals and RINGDATA-CND database (https://www.ringdata.com/news/aboutCND). The latter data source covers over 1300 major Chinese daily newspapers since 2000. In total, the sub-sample with red alerts accounted for 2.8% of the full sample.

adverse weather conditions such as fog, we conduct a sub-sample analysis by excluding foggy days from the full sample. Because fog occurs with high relative humidity, we drop days on which daily total precipitation is above 1 mm and relative humidity is above 85%. Column 6 of Table 5 shows that the coefficient estimate is statistically significantly different from 0 at the 1% level and about 33% lower than our baseline estimate. The last column of Table 5 reports the results using a subset of the original data with flight conditions recorded "good" for flight departures. That involves excluding days with heavy rainfall, high relative humidity, low visibility and flight controls due to military activities. Once again, the coefficient estimate in this column is statistically significant at the 1% level and similar in magnitude to the baseline estimate. These results in columns 6 and 7 suggest that weather confounding might result in overestimation of the true effect of pollution on flight delays.

We conduct a similar exercise to test the validity of the exclusion restriction assumption with $PM_{2.5}$ instrumented by the strength of inversions. Combined with the results in Table 5, the results of these additional analyses, reported in Appendix Table S5, strengthen our confidence in the validity of the exclusion restriction assumption in our setting.

### 4.4. *Effects of multiple air pollutants*

One challenge for estimating the adverse impacts of air pollution is that multiple air pollutants, such as $SO_2$, CO, $NO_2$, and $PM_{2.5}$ are often emitted from the same local pollution sources and may be co-transported by the wind from upwind pollution sources. Thus, concentrations of these air pollutants are often strongly correlated. The estimated impact of $PM_{2.5}$ on flight delays may be partially due to other pollutants. To probe this, we add $SO_2$, CO, $O_3$, and $NO_2$ progressively as additional controls in our preferred baseline specification. Similar to our analysis of $PM_{2.5}$, concentrations of these other pollutants are also treated as endogenous and are instrumented by wind direction.

Table 6 reports that the estimated impacts of $PM_{2.5}$ on net flight delays are still positive and statistically significant and comparable with our baseline estimate. Again, the large first stage Kleibergen-Paap F-statistics indicate that wind direction is a strong predictor of local pollution concentrations. With the inclusion of other air pollutants as controls, we estimate that a one-standard-deviation increase in daily $PM_{2.5}$ concentration is associated with an increase in net delay per flight ranging from 4.7 min to 6.3 min. The estimated impacts of $SO_2$ and $NO_2$ are neither economically nor statistically significant. The coefficient estimates for the CO variable are negative but are not statistically different from 0. Columns 4 and 5 show that $O_3$ has a puzzling negative effect. However, this finding is in line with several prior studies estimating the impacts of air pollution on infant health (Currie and Neidell, 2005), elderly mortality (Deryugina et al., 2019), and road safety (Sager, 2019). The consensus that these studies reach is that the negative correlation of $O_3$ with other air pollutants that affect human health can yield wrong-signed coefficient estimates of the $O_3$ variable. Nonetheless, the findings reported in Table 6 suggest that the estimated impacts of air pollution on flight delays are indeed due to $PM_{2.5}$ and are not related to these other pollutants.

**Table 6**
IV Estimates of effect of $PM_{2.5}$ on net flight delays when controlling for other pollutants.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $PM_{2.5}$ | 0.119*** | 0.127*** | 0.130** | 0.165*** | 0.169*** |
| | (0.0352) | (0.0402) | (0.0555) | (0.0512) | (0.0579) |
| $SO_2$ | | −0.0314 | −0.0306 | −0.0438 | −0.0388 |
| | | (0.0481) | (0.0494) | (0.0435) | (0.0469) |
| CO | | | −0.193 | −1.460 | −1.446 |
| | | | (2.291) | (2.327) | (2.311) |
| $O_3$ | | | | −0.124*** | −0.127*** |
| | | | | (0.0268) | (0.0324) |
| $NO_2$ | | | | | −0.0141 |
| | | | | | (0.0664) |
| Observations | 74,122 | 74,122 | 74,122 | 74,122 | 74,122 |
| F-test (KP statistics) | 64.91 | 29.51 | 118.9 | 119.1 | 37.00 |

Note: This table reports estimated impacts of daily average $PM_{2.5}$ and other air pollutants on daily average net delay per flight. All pollutants are instrumented by daily wind direction. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average levels of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

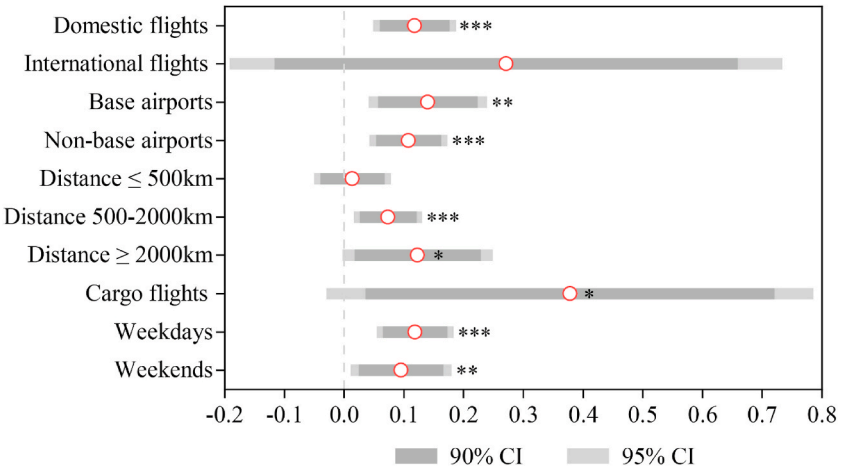



**Fig. 2.** Heterogeneity analysis. Note: This figure shows the heterogeneous effects of daily average $PM_{2.5}$ level on daily average net delay per flight, using wind direction as the instrument for $PM_{2.5}$. Each estimate is obtained by estimating equations (1) and (3) and using subsamples. The subsamples are generated as follows: (1) considering whether a flight is a domestic flight or international flight; (2) whether a flight departs from its base airport; (3) travelling distance; (4) weekdays vs. weekends; and (5) cargo flights. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Red circles indicate point estimates. The supporting regression results of the figure are presented in Appendix Table S5. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

### 4.5. *Heterogeneous effects of $PM_{2.5}$ on flight delays*

In this section, we investigate the heterogeneous effects of $PM_{2.5}$ pollution on flight delays, with wind direction as the instrument for $PM_{2.5}$. We first assess whether the estimated impact of pollution on delays differs by destination (domestic vs. international flights), because airports may give takeoff priority to international flights. We split the full sample into two subsamples and re-estimate equations (1) and (3). Holding all else the same, passengers taking domestic flights experienced an additional 4.4 min in net delay per flight for each one-standard-deviation increase in daily $PM_{2.5}$ concentrations. The pollution impact on international flights is not statistically significant.[13]

In the second investigation, we assess whether the effect differs if a flight departs from its base airport. We do so because airports may give extra assistance to flights leaving from home airports in order to prevent departure delays. The regression results displayed in Fig. 2 show the persistence of the adverse impacts of pollution on departure delays, suggesting that departing from base airports does not alleviate flight delays on polluted days.

Next, we investigate how the estimated impact of pollution on flight delays differs by travelling distance. We expect that long-distance flights would, on average, experience more severe departure delays on polluted days, relative to short-distance flights, because airport ground and cabin crews may make additional logistical preparations for long-haul flights (food, drink, etc.), resulting in relatively longer-term exposure to $PM_{2.5}$ pollution. The regression results indicate that the negative impact of pollution on flight departure delays increases with travelling distance. Specifically, the impact on flight delays is neither economically nor statistically significant when the travelling distance is less than 500 km. For flights travelling between 500 and 2000 km, each one-standard-deviation increase in the daily $PM_{2.5}$ concentration is associated with an additional 2.7 min in net departure delay per flight. The delay increases to 4.6 min for flights travelling over 2000 km. This provides suggestive support for the hypothesis that more time spent experiencing pollution affects the performance of airline crews.

We also explore how flight delays differ across types of days. The demand for air travel on weekdays is usually for business, which is likely to be fairly inelastic. On the other hand, leisure-oriented air travel is likely to occur on weekends and to be more elastic. Thus, the heterogeneity analysis for different types of days can capture whether the pollution impacts on flight delays differ by air travel demand. The regression results in the bottom of Fig. 2 show that the adverse impact of pollution on flight delays does not substantially differ for weekdays versus weekends. We also find that cargo flights experience significantly longer delays relative to passenger flights, possibly because airports give higher priority to transporting passengers than general cargo.

Finally, we examine the heterogeneous effects of $PM_{2.5}$ pollution on flight delays across our sample airports, which differ substantially by geography, operational efficiency, and systems, among other factors. We divide the full sample into five equal groups based on each airport's annual total number of passengers and estimate equations (1) and (3) for each group, with $PM_{2.5}$ instrumented by wind direction. This generates marginal airport-specific responsiveness of flight delays to a unit increase in $PM_{2.5}$. Fig. 3 reveals that

[13] The impact on international flights is not precisely estimated due to the small size of the subsample. Appendix Table S6 reports regression results for heterogeneity analyses.

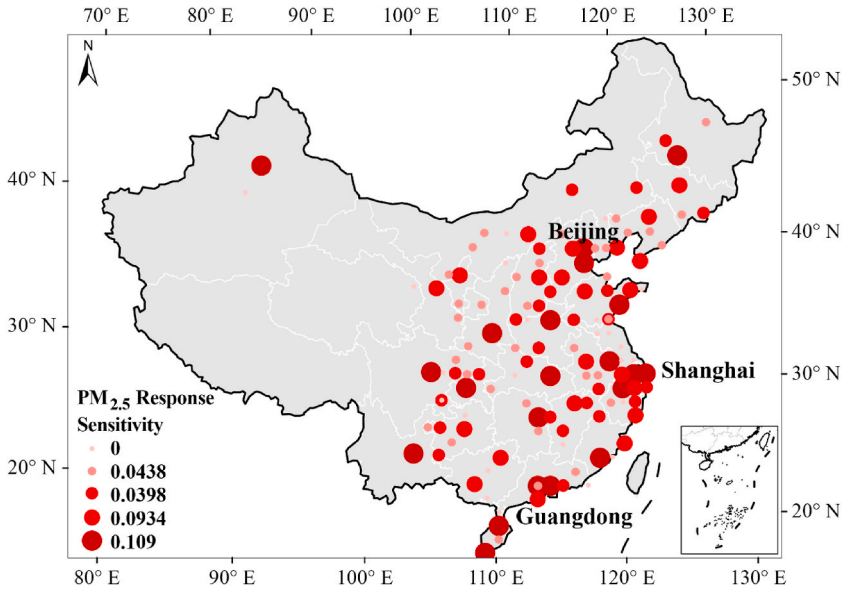



**Fig. 3.** The heterogeneous effects of $PM_{2.5}$ on net flight delays across airports. Note: The map shows the airport-specific responsiveness of net delay per flight to $PM_{2.5}$. The 89 airports covered by our sample are equally divided into five groups based on the total number of passengers carried by each airport. We obtain airport-specific responsiveness of net flight delays to the change in $PM_{2.5}$ by estimating equations (1) and (3) for these groups, with $PM_{2.5}$ instrumented by wind direction.

the adverse impacts of $PM_{2.5}$ on flight delays are more severe in large airports than those in medium-sized or small airports. Holding all else equal, for each 1 $\mu g/m^3$ increase in daily average $PM_{2.5}$ concentration, net delay per flight increases by 0.11 min in large airports, while the corresponding increases are 0.09 min in medium-sized airports and less than 0.04 min in small airports.

### *4.6. Reduced flight delays and economic benefits due to improved air quality*

China released its Air Pollution Prevention and Control Action Plan in 2013. Since then, aggressive actions have been undertaken by the Chinese government to improve air quality. Starting in 2013, the nation witnessed significant declines in $PM_{2.5}$ concentrations across the country (Zhang et al., 2019). Based on this assessment of $PM_{2.5}$ on flight delays, we analyze the contribution of air quality improvement to reductions in flight departure delays in 2015–2017.

To compute the changes in *net* (or *gross*) delay per flight for all sample airports due to improved air quality, we multiply the estimated coefficients of net (or gross) delay per flight (Column 6 of Table 2) by the levels of air quality improvement during 2015–2017 in cities where our sample airports are located. For simplicity, the levels of air quality improvement in each city are the differences in the mean annual $PM_{2.5}$ concentrations between 2015 and 2017. When calculating the changes in the national average flight delays, we multiply the same coefficient estimates by the changes in national average $PM_{2.5}$. The observed national average net delay per flight increased by 4.64 min in 2017 compared to that in 2015. However, the net delay would have increased by an additional 0.67 min if the regulation-induced improvement in air quality were absent, indicating a contribution of 13% ($\approx 0.67/(4.64 + 0.67)$). The contribution of air quality improvement was substantially larger in airports with larger passenger capacity or those located in more polluted areas. For instance, the net delay per flight at Beijing Capital Airport was reduced by 2.11 min (or 31%) due to the improvement of air quality in 2015–2017, while that at Shanghai Pudong airport, the major aviation hub of East Asia, was reduced by 1.20 min (or 22%). The contribution of air quality improvement to the reductions in gross flight delay is also significant. In 2017, the gross delay per flight at the national level increased by 4.52 min compared to that in 2015, but it would have increased by an additional 0.97 min in the absence of air quality improvement, resulting in an estimated contribution of 18% ($\approx 0.97/(4.52 + 0.97)$). The contributions of improved air quality to reductions in flight delays are also considerable at Beijing Capital (27%) and Shanghai Pudong (41%) airports.

To assess the economic benefits from improved air quality, we transform the reduction in gross flight delays due to air quality

improvement into avoided total passenger time loss. Specifically, we multiply gross delay per flight by the number of passengers carried by a flight to calculate the total passenger time loss for a flight.[14] We aggregate the total passenger time loss of all flights to obtain daily total passenger time at the airport level and then compute daily total passenger time loss per flight by dividing daily total passenger time by daily total number of flights.

To study the effect of $PM_{2.5}$ on daily total passenger time loss per flight, we estimate the same version of equations (1) and (3) using daily total passenger time loss per flight at each airport as the dependent variable. Appendix Table S7 reports our IV estimates of the causal effect of daily $PM_{2.5}$ on daily total passenger time loss per flight. We find that each 1 $\mu g/m^3$ increase in daily average $PM_{2.5}$ concentration is associated with an increase of 21.5–27.5 min of total passenger time loss per flight ($p < 0.01$). We multiply the coefficient estimates by the total number of flights at each airport in 2017 and the changes in $PM_{2.5}$ during the 2015–2017 period, summed over all sample airports, to calculate the total changes in avoided passenger time loss due to air quality improvement.

Our back-of-the-envelope calculations indicate that reduced flight delays stemming from air quality improvement avoided approximately 3.5–4.5 million hours of passenger time loss in our sample cities in 2017. If the estimated relationship between flight delays and air pollution can be extended to other flights outside the cities in our sample, the total avoided passenger time loss for all cities with airports in China would be approximately 10.7–13.6 million hours in 2017.[15]

Next, we use average income to value passenger time. The average annual income in our sample cities in 2017 was CNY 65,400, with the highest in Beijing (CNY 134,994) and the lowest in Yichun (CNY 38,713). Assuming 250 working days per year and eight working hours per day, the average income in our sample cities is equivalent to CNY 34.3 per hour. We multiply the avoided passenger time loss for each airport by city-specific hourly wages to compute the economic benefit. We estimate that the air quality improvement in 2015–2017 may have generated a gain of CNY 210–270 million (equivalent to US $32-$41 million) for our sample cities in 2017, or CNY 350–448 million (equivalent to US $54–69 million) for all cities with airports in China.

These benefit calculations are approximations. When estimating the national total avoided passenger time loss due to air quality improvement, we assume that the estimated relationship between pollution and flight delays is applicable to other flights outside of our sample, which might not be as accurate as expected. We have also used the average income to value passenger time. To the extent that business travelers earn higher income, we may have greatly underestimated the true value of passenger time and the total benefit of the avoided passenger time loss due to improved air quality. Nonetheless, our calculations suggest that the economic benefit of avoided flight delays due to air quality improvement is considerable.

## 5. Conclusions and discussion

Our analysis shows that daily variations in ground-level air pollution significantly affect flight delays. Specifically, using two separate identification approaches, we find that a one-standard-deviation increase in the $PM_{2.5}$ concentration was associated with about 6.4 min of gross delay per flight and 4.5 min of net delay per flight, which accounted for 27% and 19%, respectively, of China's national average departure delay per flight of 24 min in 2017 (Civil Aviation Administration of China, 2018). These results are robust to a wide range of variations in specifications, data and estimation strategies. Encouragingly, we also find that observed decreases in $PM_{2.5}$ concentrations attributable to China's Air Pollution Prevention and Control Action Plan avoided 13–18% of the increase in national average flight delays in 2017. We estimate that the economic benefit from avoided passenger time losses stemming from the reduction in $PM_{2.5}$ in China was about US $54-$69 million in 2017 alone.

Our findings are subject to two major limitations. First, as noted above, our sample only covers flights departing from or arriving at airports in northern China. Thus, estimated departure delays for the airports located outside of northern China are not estimated. Second, airports or airlines may have unobserved optimization or management strategies to reduce flight delays, leading to a possible underestimation of departure delays solely due to air pollution. Hence, our estimates should be interpreted cautiously as the lower bound of the true pollution impacts on flight delays.

When taken together, the robust correlation between flight departure delays and air pollution demonstrates a substantial impact of air pollution that has been neglected by current estimates of air pollution costs. Our study therefore further justifies China's adoption of stringent policies to improve air quality. The findings of our study also suggest that airports, airlines, and airport partners should take aggressive measures to reduce departure delays on polluted days, especially for passengers travelling long distances and departing from large airports. It may be necessary to enhance the air transportation system using high-tech monitoring technology, expanding the runway throughput, or increasing the number of screening officers at passenger screening checkpoints when $PM_{2.5}$ levels are high.

## Acknowledgements

Xiaoguang Chen acknowledges financial supports from the National Natural Science Foundation of China (72061147001). Wei Xie

[14] The number of passengers is calculated by multiplying aircraft seating capacities (numbers of seats per aircraft) and a loading factor. Seating capacities are typically unique for each aircraft type and published on the official website of airline companies. The CAAC raw data provides an alphanumeric-code aircraft type designator for each flight and the name of the company operating that flight. For aircraft types with sub-types (like Boeing 737NG), we use the average number of seats because the original data do not provide detailed information on aircraft sub-types for departure flights. We obtain monthly average passenger load factors in 2015–2017 in China from the WIND data database. https://www.wind.com.cn/en/data.html. Monthly average load factors are the best data available to adjust aircraft seating capacities due to the absence of flight-level load factors.

[15] As we noted in the data section, our flight data only cover the flights departing from or arriving at airports in North China.

acknowledges financial supports from the National Natural Science Foundation of China (72261147472). Other authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The authors declare no competing interests. We also thank Qinyu Deng from Beijing Normal University and Anfeng Zhu from Peking University for their assistance in producing the graphical representation of the results.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jeem.2023.102810.

**Appendix Table S1**
First stage – Effect of thermal inversions on $PM_{2.5}$ pollution

| | (1) | (2) |
|---|---|---|
| Inversion strength (°C) | 0.364*** | 0.314*** |
| | (0.0704) | (0.0499) |
| Observations | 74,122 | 74,122 |
| F-test (KP statistics) | 26.78 | 39.67 |
| Weather controls | – | YES |

Note: This table provides estimates of the effect of the strength of thermal inversions, which is defined as the aggregated temperature difference between the second layer (925 hPa, about 320 m) and the first layer (1000 hPa, about 110 m), on the daily average $PM_{2.5}$ concentrations. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.

**Appendix Table S2**
Effects of pollution and weather on flight delays

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Net delay per flight | | Gross delay per flight | |
| | 2SLS (Inversions) | 2SLS (Wind direction) | 2SLS (Inversions) | 2SLS (Wind direction) |
| $PM_{2.5}$ | 0.110*** | 0.119*** | 0.159** | 0.170** |
| | (0.0398) | (0.0352) | (0.0694) | (0.0671) |
| **Temperature Bins:** | | | | |
| <0 °C | – | – | – | – |
| | – | – | – | – |
| 0–5 °C | −2.775*** | −2.863*** | −5.621*** | −5.741*** |
| | (0.800) | (0.827) | (1.459) | (1.479) |
| 5–10 °C | −3.415*** | −3.477*** | −7.121*** | −7.206*** |
| | (0.924) | (0.929) | (1.610) | (1.631) |
| 10–15 °C | −3.784*** | −3.849*** | −8.927*** | −9.016*** |
| | (1.121) | (1.094) | (1.955) | (1.927) |
| 15–20 °C | −5.180*** | −5.280*** | −10.17*** | −10.31*** |
| | (1.400) | (1.296) | (2.479) | (2.388) |
| 20–25 °C | −6.532*** | −6.674*** | −12.14*** | −12.34*** |
| | (1.763) | (1.587) | (3.138) | (2.959) |
| 25–30 °C | −5.989*** | −6.197*** | −11.59*** | −11.88*** |
| | (2.192) | (1.915) | (3.815) | (3.470) |
| 30–35 °C | −2.198 | −2.442 | −5.674 | −6.007 |
| | (2.433) | (2.260) | (4.571) | (4.220) |
| >35 °C | −16.93*** | −17.26*** | −14.46* | −14.92* |
| | (5.331) | (4.892) | (8.153) | (7.728) |
| **Precipitation decile bins:** | | | | |
| Zero Precipitation | – | – | – | – |
| | – | – | – | – |
| Bottom 10% | 1.455*** | 1.505*** | 0.859 | 0.927 |
| | (0.544) | (0.498) | (0.996) | (0.891) |
| 10–20 Percentile | 0.690 | 0.744 | −0.0144 | 0.0593 |
| | (0.457) | (0.512) | (0.856) | (0.887) |
| 20–30 Percentile | 1.249** | 1.307** | 0.515 | 0.594 |
| | (0.623) | (0.618) | (0.995) | (0.964) |
| 30–40 Percentile | 1.123* | 1.195* | 0.343 | 0.442 |

(continued on next page)

**Appendix Table S2** (*continued*)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Net delay per flight | | Gross delay per flight | |
| | 2SLS (Inversions) | 2SLS (Wind direction) | 2SLS (Inversions) | 2SLS (Wind direction) |
| | (0.669) | (0.608) | (1.035) | (0.955) |
| 40–50 Percentile | 1.485** | 1.554** | 1.615 | 1.709 |
| | (0.622) | (0.599) | (1.157) | (1.053) |
| 50–60 Percentile | 2.970*** | 3.055*** | 3.461*** | 3.577*** |
| | (0.743) | (0.736) | (1.298) | (1.220) |
| 60–70 Percentile | 3.026*** | 3.117*** | 2.884** | 3.010** |
| | (0.716) | (0.695) | (1.330) | (1.217) |
| 70–80 Percentile | 3.286*** | 3.382*** | 4.040*** | 4.173*** |
| | (0.874) | (0.878) | (1.441) | (1.351) |
| 80–90 Percentile | 5.040*** | 5.133*** | 6.128*** | 6.256*** |
| | (1.098) | (1.008) | (1.798) | (1.604) |
| Top 10 Percentile | 9.442*** | 9.546*** | 14.34*** | 14.49*** |
| | (1.479) | (1.322) | (2.526) | (2.259) |
| Wind Speed | 0.693*** | 0.729*** | 1.657*** | 1.707*** |
| | (0.238) | (0.232) | (0.440) | (0.445) |
| Water Vapor Pressure | 0.237** | 0.240*** | 0.435** | 0.440*** |
| | (0.0923) | (0.0856) | (0.173) | (0.167) |
| Air Pressure | 0.00851 | 0.00988 | 0.000664 | 0.00254 |
| | (0.0149) | (0.0158) | (0.0258) | (0.0273) |
| Relative Humidity | −0.00939 | −0.0129 | 0.0596 | 0.0547 |
| | (0.0271) | (0.0196) | (0.0471) | (0.0369) |
| Observations | 74,122 | 74,122 | 74,122 | 74,122 |
| F-test (KP Statistics) | 39.67 | 64.91 | 39.67 | 64.91 |

Note: This table reports the estimated impacts of daily average $PM_{2.5}$ level and weather on daily average net delay per flight and gross delay per flight. Columns 1 and 3 report results using the strength of thermal inversions (aggregated temperature difference between the second and the first layers) as the instrument variable. Columns 2 and 4 report the results using daily wind direction as the instrument variables. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.

**Appendix Table S3**
Robustness – Results with additional leads of $PM_{2.5}$

| | (1) | (2) | (3) |
|---|---|---|---|
| | 2SLS (Wind Direction) | 2SLS (Wind Direction) | 2SLS (Wind Direction) |
| $PM_{2.5}$ | 0.119*** | 0.120*** | 0.0995*** |
| | (0.0352) | (0.0401) | (0.0267) |
| $PM_{2.5}$ one day from now | | −0.00726 | |
| | | (0.0229) | |
| $PM_{2.5}$ three days from now | | | 0.0297 |
| | | | (0.0263) |
| Observations | 74,122 | 70,228 | 70,480 |
| F-test (KP statistics) | 64.91 | 8.857 | 2.430 |

Note: This table reports the estimated impacts of daily average $PM_{2.5}$ on daily average net delay per flight, with additional leads of $PM_{2.5}$ instrumented by corresponding leads of daily wind directions. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, relative humidity, and corresponding leads of weather controls. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level.
***p < 0.01, **p < 0.05, *p < 0.1.

**Appendix Table S4**
Robustness of IV estimates using inversions as an instrument

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Dummy indicator of inversion as instrument | Frequency of inversion as instrument | Adding province × year FE | Adding year × month FE | Clustering SD within airports and dates | IDW within 25 km | IDW within 100 km | Removing flights between 12 a.m. and 6 a.m. | Controlling $PM_{2.5}$ at landing airports | PPML |
| $PM_{2.5}$ (departure airports) | 0.0893* | 0.0714* | 0.118*** | 0.110*** | 0.110*** | 0.117** | 0.106*** | 0.110*** | 0.0585* | 0.00676*** |
| | (0.0492) | (0.0373) | (0.0388) | (0.0393) | (0.0397) | (0.0487) | (0.0356) | (0.0398) | (0.0306) | (0.00245) |
| | | | | | | | | | 0.00540 | |

(*continued on next page*)

**Appendix Table S4** (*continued*)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Dummy indicator of inversion as instrument | Frequency of inversion as instrument | Adding province × year FE | Adding year × month FE | Clustering SD within airports and dates | IDW within 25 km | IDW within 100 km | Removing flights between 12 a.m. and 6 a.m. | Controlling $PM_{2.5}$ at landing airports | PPML |
| $PM_{2.5}$ (landing airports) | | | | | | | | | (0.00371) | |
| Observations | 74,122 | 74,122 | 74,122 | 74,122 | 74,122 | 49,160 | 86,608 | 74,122 | 436,743 | 74,122 |
| First stage F-test (KP statistics) | 39.03 | 38.01 | 40.47 | 38.92 | 37.41 | 21.12 | 50.80 | 39.67 | 21.87 | – |

Note: This table reports estimated impacts of daily average $PM_{2.5}$ on daily average net delay per flight with alternative model specifications, data and estimation strategies, using thermal inversion as the instrument. Columns 1–2 report the IV estimates using a dummy indicator of thermal inversions and the frequency of thermal inversions (count of hours with thermal-inversion events), respectively, as the instrument variable. Columns 3–10 use the strength of thermal inversions as the instrument. Columns 3-4 additionally include province × year and year × month FE, respectively. Column 5 clusters the standard errors at the airport and date levels to further address concerns on potential network effects of flight delays across regions. Columns 6–7 use a radius of 25 km and 100 km, respectively, to generate distance-weighted pollution and weather data for sample airports. Column 8 excludes overnight flights between 12 a.m. and 6 a.m. in the sample. Column 9 reports estimates from the bilateral-route specification which includes the daily average $PM_{2.5}$ level and weather controls at both the departing and landing airports. Column 10 uses a PPML specification and bootstraps standard errors with 1000 replications, while the endogeneity of $PM_{2.5}$ is addressed by using control function method. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level except in columns 5 and 10. ***p < 0.01, **p < 0.05, *p < 0.1.

**Appendix**

Table S5**Robustness of IV estimates to weather controls and weather conditions with inversions as an instrument**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Flexible weather controls for temperature and precipitation | Flexible weather controls for temperature, precipitation, and wind speed | Flexible weather controls for temperature, precipitation, and relative humidity | Adding visibility as an additional control | Excluding "red-alert" days | Excluding foggy days | Flight conditions recorded "good" for departures |
| $PM_{2.5}$ | 0.119*** | 0.121*** | 0.0963*** | 0.150*** | 0.119** | 0.0575** | 0.0825** |
| | (0.0403) | (0.0425) | (0.0365) | (0.0526) | (0.0480) | (0.0272) | (0.0338) |
| Observations | 74,122 | 74,122 | 74,122 | 74,122 | 72,061 | 51,215 | 72,575 |
| F-test (KP statistics) | 41.99 | 44.14 | 59.67 | 26.59 | 34.72 | 45.39 | 38.92 |

Note: This table reports estimated impacts of daily average $PM_{2.5}$ level on daily average net delay per flight with alternative weather controls and sub-samples, using the strength of thermal inversion as the instrument for $PM_{2.5}$. Columns 1–3 show the results equivalent to our baseline specification, but with more flexible weather controls. Column 1 includes all possible interactions for temperature (5 °C bins) and precipitation (decile bins). Column 2 includes all possible interactions for temperature (5 °C bins), precipitation (decile bins), and wind speed (decile bins). Column 3 includes all possible interactions for temperature (5 °C bins), precipitation (decile bins), and relative humidity (decile bins). Column 4 includes daily average airport-level atmospheric visibility as an additional control. Column 5 reports the estimate from a sub-sample that excludes days with red alerts. Column 6 reports the estimate from a sub-sample of days with daily total precipitation below 1 mm, relative humidity below 85%. Column 7 reports the estimate from a sub-sample when flight conditions were recorded "good" for flight departures, which excludes days with heavy rainfall, high relative humidity, low visibility, and flight controls due to military activities. Weather controls for columns 3–5 include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.

**Appendix Table S6**
Heterogeneity by subsets of sample

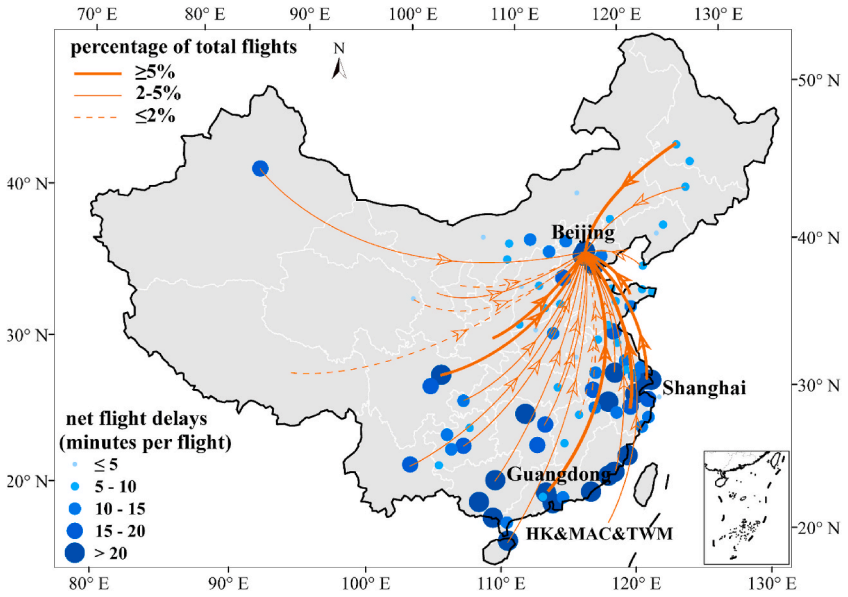
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Domestic Flights | International Flights | Flights departing from base airports | Flights departing from non-base airports | Travel distance (≤ 500 km) | Travel distance (500–2000 km) | Travel distance (≥ 2000 km) | Cargo Flights | Weekdays | Weekends |
| $PM_{2.5}$ | 0.118*** | 0.271 | 0.140** | 0.108*** | 0.0138 | 0.0735** | 0.123* | 0.378* | 0.119*** | 0.0952** |
| | (0.0355) | (0.236) | (0.0506) | (0.0333) | (0.0328) | (0.0290) | (0.0642) | (0.208) | (0.0329) | (0.0431) |
| Observations | 74,121 | 4139 | 16,853 | 74,113 | 21,287 | 108,113 | 14,204 | 9308 | 53,001 | 21,119 |
| F-test (KP statistics) | 66.33 | 0.767 | 14.96 | 64.49 | 7.690 | 72.53 | 93.43 | 90.65 | 66.26 | 18.56 |

Note: This table reports estimated impacts of daily average $PM_{2.5}$ level on daily average net delay per flight using subsets of domestic/international flights (columns 1–2), subsets of flights departing from base/non-base airports (columns 3–4), subsets of flights with different travelling distances (columns 5–7), subset of cargo flights (column 8) and subsets of weekdays and weekends (columns 9–10). Wind direction is used as the instrument for $PM_{2.5}$. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.

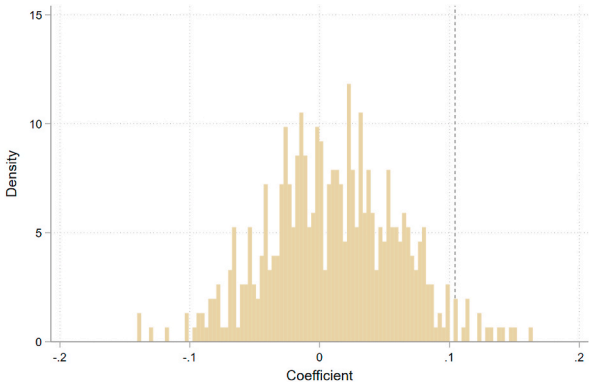
**Appendix Table S7**
Estimates of the causal effects of $PM_{2.5}$ on total passenger time loss per flight

| | (1) | (2) |
|---|---|---|
| | Total passenger time loss per flight (Wind Direction) | Total passenger time loss per flight (Strength of Thermal Inversions) |
| $PM_{2.5}$ | 21.54*** | 27.51*** |
| | (7.628) | (10.24) |
| Observations | 74,122 | 74,122 |
| F-test (KP statistics) | 66.34 | 39.67 |

Note: This table reports the estimated impacts of daily average $PM_{2.5}$ level on daily total passenger time loss per flight, which is computed using gross delay per flight. In column 1, $PM_{2.5}$ is instrumented using wind direction. In column 2, $PM_{2.5}$ is instrumented using strength of thermal inversions. Weather controls include 9 temperature bins (<0, 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, >35 °C), decile bins for precipitation, daily average level of wind speed, water vapor pressure, air pressure, and relative humidity. We also control for a set of location-specific and temporal fixed effects, including airport, day-of-week, year and month fixed effects. Standard errors are clustered at the airport and region × date level. ***p < 0.01, **p < 0.05, *p < 0.1.




**Appendix Fig. S1.** The percentage of flight takeoffs from the rest of China and foreign countries to North China and daily average net flight delays for airports covered by our sample




**Appendix Fig. S2.** Distribution of estimated coefficients from the permutation placebo tests
Note: This figure reports IV estimates of our baseline specification when using randomly generated placebo instruments. Specifically, we randomly assign the daily wind direction for each airport and construct placebo treatment analysis for the sample of individual airports as that in the baseline analysis. These histograms display the distribution of placebo treatment effects from 500 random assignments. The dashed line shows the estimated treatment from the baseline analysis. The *p* value of the permutation test is the proportion of placebo estimates that are equal to or larger in absolute value than the corresponding estimate from the baseline analysis.

## References


AlKheder, S., AlKandari, A., 2020. The impact of dust on Kuwait International Airport operations: a case study. Int. J. Environ. Sci. Technol. https://doi.org/10.1007/s13762-020-02710-3.

Arceo, E., Hanna, R., Oliva, P., 2016. Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico city. Econ. J. https://doi.org/10.1111/ecoj.12273.

Baddock, M.C., Strong, C.L., Murray, P.S., McTainsh, G.H., 2013. Aeolian dust as a transport hazard. Atmos. Environ. https://doi.org/10.1016/j.atmosenv.2013.01.042.

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A., Zou, B., Britto, R., Fearing, D., Swaroop, P., Uman, N., Vaze, V., Voltes, A.. Total delay impact study. https://isr.umd.edu/NEXTOR/pubs/TDI_Report_Final_11_03_10.pdf.

Bondy, M., Roth, S., Sager, L., 2020. Crime is in the air: the contemporaneous relationship between air pollution and crime. J. Assoc. Environ. Resour. Econ. 7, 555–585. https://doi.org/10.2139/ssrn.3170281.

Borsky, S., Unterberger, C., 2019. Bad weather and flight delays: the impact of sudden and slow onset weather events. Econ. Transp. 18, 10–26. https://doi.org/10.1016/j.ecotra.2019.02.002.

Brueckner, J.K., 2005. Internalization of airport congestion: a network analysis. Int. J. Ind. Organ. 23, 599–614. https://doi.org/10.1016/j.ijindorg.2005.03.007.

Carneiro, J., Cole, M.A., Strobl, E., 2021. The effects of air pollution on students' cognitive performance: evidence from brazilian university entrance tests. J. Assoc. Environ. Resour. Econ. 8, 1051–1077. https://doi.org/10.1086/714671.

Chay, K.Y., Greenstone, M., 2003. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. Q. J. Econ. 118, 1121–1167. https://doi.org/10.1162/00335530360698513.

Chen, X., Ye, J., 2019. When the wind blows: spatial spillover effects of urban air pollution in China. J. Environ. Plann. Manag. 62, 1359–1376. https://doi.org/10.1080/09640568.2018.1496071.

Chen, Y., Ebenstein, A., Greenstone, M., Li, H., 2013. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. Proc. Natl. Acad. Sci. USA 110, 12936–12941. https://doi.org/10.1073/pnas.1300018110.

Civil Aviation Administration of China, 2018. Statistical Bulletin on the Development of Civil Aviation Industry in 2017. http://www.caac.gov.cn/en/HYYJ/NDBG/201810/W020181026601069968468.pdf.

Correia, S., Guimarães, P., Zylkin, T.Z., 2020. Fast Poisson estimation with high-dimensional fixed effects. STATA J. 20, 95–115. https://doi.org/10.1177/1536867X20909691.

Currie, J., Neidell, M., 2005. Air pollution and infant health: what can we learn from California's recent experience? Q. J. Econ. 120, 1003–1030. https://doi.org/10.1093/qje/120.3.1003.

Deryugina, T., Heutel, G., Miller, N.H., Molitor, D., Reif, J., 2019. The mortality and medical costs of air pollution: evidence from changes in wind direction. Am. Econ. Rev. 109, 4178–4219. https://doi.org/10.1257/aer.20180279.

Deshpande, V., Arıkan, M., 2012. The impact of airline flight schedules on flight delays. Manuf. Serv. Oper. Manag. https://doi.org/10.1287/msom.1120.0379.

Ebenstein, A., Lavy, V., Roth, S., 2016. The long-run economic consequences of high- stakes examinations: evidence from transitory variation in pollution. Am. Econ. J. Appl. Econ. 8, 36–65. https://doi.org/10.1257/app.20150213.

Forbes, S.J., 2008. The effect of air traffic delays on airline prices. Int. J. Ind. Organ. 26, 1218–1232. https://doi.org/10.1016/j.ijindorg.2007.12.004.

Fu, S., Viard, V.B., Zhang, P., 2021. Air pollution and manufacturing firm productivity: nationwide estimates for China. Econ. J. 131, 3241–3273. https://doi.org/10.1093/ej/ueab033.

Gao, H., Shi, J., Cheng, H., Zhang, Yaqin, Zhang, Yan, 2021. The impact of long- and short-term exposure to different ambient air pollutants on cognitive function in China. Environ. Int. 151, 106416 https://doi.org/10.1016/j.envint.2021.106416.

Ghanem, D., Zhang, J., 2014. Effortless Perfection:' Do Chinese cities manipulate air pollution data? J. Environ. Econ. Manag. 68, 203–225. https://doi.org/10.1016/j.jeem.2014.05.003.

Greenstone, M., He, G., Jia, R., Liu, T., 2022. Can technology solve the principal-agent problem? Evidence from China's war on air pollution. Am. Econ. Rev. Insights 4, 54–70. https://doi.org/10.1257/aeri.20200373.

Gultepe, I., Tardif, R., Michaelides, S.C., Cermak, J., Bott, A., Bendix, J., Müller, M.D., Pagowski, M., Hansen, B., Ellrod, G., Jacobs, W., Toth, G., Cober, S.G., 2007. Fog research: a review of past achievements and future perspectives. Pure Appl. Geophys. 164, 1121–1159. https://doi.org/10.1007/s00024-007-0211-x.

Hanna, R., Oliva, P., 2015. The effect of pollution on labor supply: evidence from a natural experiment in Mexico City. J. Publ. Econ. 122, 68–79. https://doi.org/10.1016/j.jpubeco.2014.10.004.

He, J., Liu, H., Salvo, A., 2019. Severe air pollution and labor productivity: evidence from industrial towns in China. Am. Econ. J. Appl. Econ. 11, 173–201. https://doi.org/10.1257/app.20170286.

Hyslop, N.P., 2009. Impaired visibility: the air pollution people see. Atmos. Environ. https://doi.org/10.1016/j.atmosenv.2008.09.067.
Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–475.
Jans, J., Johansson, P., Nilsson, J.P., 2018. Economic status, air quality, and child health: evidence from inversion episodes. J. Health Econ. https://doi.org/10.1016/j.jhealeco.2018.08.002.
Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525, 367–371. https://doi.org/10.1038/nature15371.
Mayer, C., Sinai, T., 2003. Network effects, congestion externalities, and air traffic delays: or why not all delays are evil. Am. Econ. Rev. 93, 1194–1215. https://doi.org/10.1257/000282803769206269.
Mazzeo, M.J., 2003. Competition and service quality in the U.S. airline industry. Rev. Ind. Organ. 22, 275–296. https://doi.org/10.1023/A:1025565122721.
Rupp, N.G., Holmes, G.M., 2006. An investigation into the determinants of flight cancellations. Economica 73, 749–783.
Sager, L., 2019. Estimating the effect of air pollution on road safety using atmospheric temperature inversions. J. Environ. Econ. Manag. 98, 1–20. https://doi.org/10.1016/j.jeem.2019.102250.
Santos, G., Robin, M., 2010. Determinants of delays at European airports. Transp. Res. Part B Methodol. https://doi.org/10.1016/j.trb.2009.10.007.
Schlenker, W., Walker, W.R., 2016. Airports, air pollution, and contemporaneous health. Rev. Econ. Stud. 83, 768–809. https://doi.org/10.1093/restud/rdv043.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586. https://doi.org/10.2307/2171753.
Tang, D., Hoshiko, E., 2013. China Pollution Grounds Hundreds of Flights, Prompts Severe Health Warning. NBC NEWS.
The World Bank, 2018. Air transport, passengers carried-China [WWW Document]. URL. https://data.worldbank.org/indicator/IS.AIR.PSGR?end=2018&locations=CN&start=1998&view=map.
Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang, X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., Yan, L., Geng, G., Hong, C., Li, M., Liu, F., Zheng, B., Cao, J., Ding, A., Gao, J., Fu, Q., Huo, J., Liu, B., Liu, Z., Yang, F., He, K., Hao, J., 2019. Drivers of improved PM2.5 air quality in China from 2013 to 2017. Proc. Natl. Acad. Sci. U.S.A. https://doi.org/10.1073/pnas.1907956116.
Zhang, Xin, Chen, X., Zhang, Xiaobo, 2018. The impact of exposure to air pollution on cognitive performance. Proc. Natl. Acad. Sci. USA 115, 9193–9197. https://doi.org/10.1073/pnas.1809474115.
Zhang, Xin, Zhang, Xiaobo, Chen, X., 2017. Happiness in the air: how does a dirty sky affect mental health and subjective well-being? J. Environ. Econ. Manag. 85, 81–94. https://doi.org/10.1016/j.jeem.2017.04.001.
Zhao, T., Markevych, I., Romanos, M., Nowak, D., Heinrich, J., 2018. Ambient ozone exposure and mental health: a systematic review of epidemiological studies. Environ. Res. 165, 459–472. https://doi.org/10.1016/j.envres.2018.04.015.
Zheng, S., Wang, J., Sun, C., Zhang, X., Kahn, M.E., 2019. Air pollution lowers Chinese urbanites' expressed happiness on social media. Nat. Human Behav. 3, 237–243. https://doi.org/10.1038/s41562-018-0521-2.
Zhuang, P., 2016. Hundreds of flights cancelled in Beijing as thick smog lays siege to capital. South China Morning Post. https://www.scmp.com/news/china/society/article/2056009/hundreds-flights-cancelled-beijing-thick-smog-lays-siege-capital.
Zivin, J.G., Neidell, M., 2012. The impact of pollution on worker productivity. Am. Econ. Rev. 102, 3652–3673. https://doi.org/10.1257/aer.102.7.3652.