Journal of Development Economics 171 (2024) 103340

Contents lists available at ScienceDirect

Journal of Development Economics

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

Regular Article

Pollution-induced trips: Evidence from flight and train bookings in China☆

Ruochen Dai[a], Dongmei Guo[a], Yajie Han[b], Yu Qin[b,*]

[a] Central University of Finance and Economics, China
[b] NUS Business School, National University of Singapore, Singapore

ARTICLE INFO

JEL classification:
O15
Q53
Q56
R40

Keywords:
Air pollution
Travel flow
Avoidance behavior

ABSTRACT

Utilizing a novel database including nearly 2.2 billion booking records in China, we examine whether people escape from pollution by traveling to "cleaner" places. Combining an instrumental variable approach with high-dimensional fixed effects, we find a 50-unit increase in the AQI gap between a city pair leads to a 1.30% (1.33%) increase in train and airline ticket bookings from the origin to the destination city departing within one day (2–7 days). In addition, the destination of such pollution-induced trips is more likely to be an intra-province city with more tourist attractions. We also measure willingness to pay for clean air.

1. Introduction

In recent years, with air-pollution disclosure and publicity about the importance of air quality, people have become more aware of air-pollution issues (Barwick et al., 2024). To reduce the potential adverse impacts of air pollution on health, productivity, and cognition (Chang et al., 2019; Chen et al., 2013; Currie and Neidell, 2005; Ebenstein et al., 2016; Fan et al., 2020; Guarnieri and Balmes, 2014; He et al., 2019; Herrnstadt et al., 2021; Knittel et al., 2016; Landrigan et al., 2018; Zhang et al., 2018), people may take different strategies to avoid exposure to air pollution, such as staying indoors, purchasing air purifiers, wearing masks, and migrating to cities or countries with cleaner air (Ito and Zhang, 2020; Khanna et al., 2021; Qin and Zhu, 2018; Sun et al., 2019; Zhang and Mu, 2018). However, the avoidance behaviors mentioned above fail in some cases. First, reducing indoor pollution by air purifiers and avoiding outdoor activities may not be optimal solutions in some cases—outdoor activities are essential for physical and mental health (Jung et al., 2019; Triguero-Mas et al., 2017). In addition, permanent migration, as another way to avoid air pollution, is a costly decision (Freeman et al., 2019). Considering the high cost of permanent migration and the relatively low frequency of extremely polluted days, short-term trips can be a complementary option to avoid air pollution.

In this paper, we investigate how air pollution affects people's travel decisions in the short run. Utilizing a novel database from one of the largest flight and train bookings providers in China, which includes nearly 2.2 billion booking records with detailed individual and price information from 2017 to 2019, we examine whether people book trips to places with relatively cleaner air when air pollution in their residential places becomes severe. We construct the travel flow by train and air between each city pair in the perspective of the reservation date of the bookings and link the travel booking records with the air quality of the origin and destination city on the reservation date. Based on this data and the total trips of active users of the booking platform, we measure the *probability* of individuals from an origin city choosing to travel to a destination city on a given reservation date. We could match the travel decisions to the air pollution on the exact location and *date*

☆ We are grateful to the editor and the two anonymous referees for their insightful suggestions. We also thank Shiqi Guo, Wei Shi, Min Wang, Jianwei Xing, Jubo Yan, as well as participants at the 2021 CCER Summer Institute, the 15th North American Meeting of the Urban Economics Association, the 2021 Jinan-SMU-ABFER Conference on Urban and Regional Economics, the 2022 Asian Meeting of the Econometric Society, the 2023 PKU Summer Meeting in Urban Economics, and seminar participants at the National University of Singapore, Peking University, and Nanyang Technological University, for their valuable feedback. We gratefully acknowledge the financial support from the Major Program of the National Social Science Foundation of China (21&ZD085), the National Natural Science Foundation of China (72342031, 72203252), and the Program for Innovation Research at the Central University of Finance and Economics. We also appreciate Yanjun Xu and Tao Wang for their excellent assistance in data processing as research consultants. All analyses and data sourcing were conducted in the People's Republic of China. All errors are ours.

* Corresponding author.
E-mail addresses: r.dai@cufe.edu.cn (R. Dai), guodongmeicufe@163.com (D. Guo), yajie.h@u.nus.edu (Y. Han), bizqyu@nus.edu.sg (Y. Qin).

https://doi.org/10.1016/j.jdeveco.2024.103340
Received 6 February 2023; Received in revised form 2 June 2024; Accepted 9 July 2024
Available online 18 July 2024
0304-3878/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

that the travel decisions are made.

To address the potential endogeneity of air quality affected by human activities, we use upwind pollution based on wind directions and wind speeds as an instrumental variable (IV) for the air-quality gap between the origin and destination cities. In addition, we also control for a rich set of fixed effects, including city-pair fixed effects (to control for city-pair-level time-invariant unobservables), date fixed effects (to control for a flexible time trend), and origin-year-month and destination-year-month fixed effects to allow for a flexible time trend for each origin and destination city-pair. Finally, we also control for weather differences for each city-pair date that may affect city-pair travel flows, including differences in wind speed, temperature, precipitation, sunshine, relative humidity, and atmospheric pressure.

Our analyses reveal four main findings. First, we find a 50-unit increase in the Air Quality Index (AQI) gap between the origin city and the destination city leads to a 1.30% and 1.33% significant increase in travel bookings, departing within one day and two to seven days respectively. Such effects are slightly larger for train ticket bookings than airline ticket bookings. As a falsification test, we also show the AQI gap on the reservation day has no impact on travel bookings that depart in more than one week. In addition, the results are similar if we use an alternative measure of air pollution ($PM_{2.5}$) or alternative standard errors (one-way clustering at the city-pair level instead of two-way clustering at the origin- and destination-city level). Additionally, the increased train bookings on more polluted days are newly generated by air pollution, whereas flight bookings are more likely to be the intertemporal shifting of trips planned previously.

Second, we investigate the heterogeneity of the results. We show the effect is the largest for people ages 19–35 and becomes smaller in older cohorts. In addition, we find the effect is largely driven by intra-province travels instead of inter-province travels. In particular, the results show that on polluted days, people buy train tickets to cleaner cities with tourist attractions within the same province, which supports the conjecture that the purpose of the trips reserved on polluted days is to escape bad air quality. In addition, we find that the effects of AQI difference between a city pair are more prominent if the origin city is on average a high-pollution city, and if the destination city is on average a clean city.

Third, by exploiting the refund and cancellation of train and flight bookings, we show train bookings on days with a higher AQI gap lead to a higher probability of refund afterward, suggesting individuals are more likely to regret their train bookings made on high-AQI-gap days. The above results suggest that in addition to rational explanations for our findings, such as pollution avoidance, behavioral factors may play a role, such as projection bias.

Lastly, we measure willingness to pay (WTP) for clean air by running a reduced-form analysis and utilizing the price information in our data. Our OLS estimation suggests people leaving within one day (2–7 days) are willing to pay 0.14 (0.31) yuan for a one-unit improvement of AQI. However, OLS would underestimate the coefficient of AQI difference and overestimate the coefficient of price in the reduced-form analysis. By instrumenting for both price and AQI difference, we estimate that an individual who departs within one day (2–7 days), on average, is willing to pay 0.31 (0.10) yuan for a one-unit improvement of AQI. In other words, an individual would be willing to pay 0.31 yuan for each unit improvement in AQI within one day, and 0.10 yuan for each unit improvement in AQI for avoiding pollution within 2–7 days. The IV estimate is larger than the OLS estimate for trips leaving within one day because air pollution difference suffers a more severe reverse-causality problem for trips leaving within one day.[1]

Our paper contributes to the large literature investigating pollution-avoidance behavior, including staying indoors (Guarnieri and Balmes, 2014; Sun et al., 2019), purchasing masks (Zhang and Mu, 2018), and using air filters (Ito and Zhang, 2020), by showing a short-term trip is an alternative option to avoid air pollution, and the air quality on the reservation date does affect travel decisions. Our paper is most related to a few recent works examining air pollution and travel flows (Barwick et al., 2021; Chen et al., 2020, 2021).[2] Our paper differs from those three papers in the following aspects. First, instead of linking travel flows to the air pollution on the travel dates, we are the first paper to link the travel flows to the air pollution on the *dates that the travel decisions are made*, that is, the reservation dates. Linking air pollution to the decision date of travels is important because pollution-induced travels may not take place on the same day. Second, our unique data from the largest travel booking platform in China allow us to examine rich heterogeneities of the effects, including bookings by travel mode, by age cohort, and by city characteristics. Third, our data on ticket refund and cancellation allow us to investigate the existence of behavioral factors. Lastly, the price information in our data allows us to directly measure the WTP for clean air.

Our paper is also related to the literature estimating WTP for clean air. Ito and Zhang (2020) find a household is willing to pay \$1.34 (about 9 yuan) annually to remove 1 $\mu g/m^3$ of PM_{10}, and Chen et al. (2021) find the average marginal WTP for AQI is around 6.4 yuan. Our estimates (2.24–3.36 and 0.72–1.08 yuan)[3] are smaller than the WTP estimated from air purifier purchases by Ito and Zhang (2020) and much smaller than the WTP estimated from long-term migration by Bayer et al. (2009) and Freeman et al. (2019), which is reasonable because air purifiers are durable goods and migration is a long-term decision, whereas travel is transient consumption. However, our estimate is likely to be a lower bound because we do not consider other costs associated with such short-term travels, such as accommodation and transportation costs in the destination cities. Our work shows another dimension of measuring WTP for air quality, which could complement the existing literature.

Our paper provides important implications for environmental policies. We show air pollution induces more travel, which, on the one hand, brings health benefits to the travelers. However, on the other hand, such pollution-induced travels offset the efforts in greenhouse gas emission (GHG) reductions. According to our back-of-the-envelope calculation, the increased travel due to air pollution leads to an 8,292.45-ton increase in CO2e annually, which means we need to plant 460,692–2,073,114 more trees to absorb these additional GHG emissions.

The paper proceeds as follows. Section 2 describes the data used. Section 3 introduces the empirical strategy including the construction of the instrumental variable. Section 4 presents the estimation results. Section 5 discusses the irrational aspect of this avoidance behavior, intertemporal shifting, and the indirect cost of the pollution-induced trips and Section 6 concludes.

[1] We get similar estimates of WTP using $PM_{2.5}$ as the measure of air pollution. Specifically, an individual is willing to pay 0.31 yuan for each unit improvement in $PM_{2.5}$ for avoidance of 1-unit $PM_{2.5}$ within one day, and 0.11 yuan for each unit improvement in $PM_{2.5}$ for avoiding pollution within 2–7 days.

[2] Specifically, Chen et al. (2021) find the increase in the air-pollution difference between origin and destination cities increases short-term population flow from origin to destination, using cell phone data and Beijing airport flights data, respectively. Utilizing credit- and debit-card transaction data, Barwick et al. (2021) show high-speed railways and air-travel networks facilitate intercity travel, which acts as an effective means of adaption to air pollution.

[3] To make the WTP comparable to those presented in Ito and Zhang (2020), we make an annualization of our WTP estimates. An individual who departs within one day (2–7 days), on average, is willing to pay 0.31 (0.10) yuan for a one-unit improvement in AQI. In our data, an individual travels 3.61 times per year, on average, which means an average individual leaving in zero to one days (2–7 days) is willing to pay 1.12 (0.36) yuan annually for a one-unit improvement in AQI. Therefore, the WTP of a household of two to three is about 2.24–3.36 (0.72–1.08) yuan for one unit of AQI in a trip leaving in zero to one days (2–7 days), respectively.

2. Empirical background and data

2.1. Online transportation ticket bookings

With the rapid development of the internet, smartphones, and mobile payments, online ticket booking has become more popular due to its speed and cost-effectiveness. As of June 2019, the number of online travel booking users in China reached 418 million, accounting for 48.9% of the total netizens.[4] In 2018, people who booked train (airline) tickets from the ticket bookings provider we acquire data from (hereinafter referred to as "the institution" or "the ticket bookings provider") accounted for 16.2% (22.8%) of people who traveled by train (air) throughout the year.[5]

In China, transportation tickets can be booked online in two primary ways. One way is to book on official websites: airline tickets on the airline's websites such as *China Southern Airlines*, and train tickets on China Railway Customer Service Center, known as *12306 China Railway*.[6] The other way is to book using online travel agencies (OTA) such as Ctrip, Fliggy, Qunar, and Meituan, where people can make reservations on their chosen flight, train, and bus routes through the mobile platform, internet websites, and customer service centers and arrange electronic payment.

2.2. Ticket bookings data

We utilize anonymized ticket booking data obtained from one of the largest online travel agencies (OTA) in China, which is also one of the largest travel service providers in the world. Its main services include accommodation reservations, train and airline ticket bookings, and packaged tours.

The online travel agency industry in China has enjoyed prosperous growth in the past decade, mainly thanks to the popularity of smartphones. The total revenue reached 1.675 trillion USD in 2019 with an online penetration rate of around 20%.[7] The professional institution we get data from was founded in 1999 as a pioneer in this industry and has become a travel agency empire in the past 20 years. The institution acquired the second-largest OTA at that time in 2015, and has since controlled around half of the market. It acquired a large train-ticket-booking platform in 2001 and became the largest third-party agency in the online train-ticket-booking market. In particular, it is widely preferred by individuals and corporations who want to book tickets online, covering more than a half of the online travel ticket booking market in 2018.[8] Overall, the provider is regarded as the most powerful OTA in China, and the data we use are the most representative data for analyzing Chinese people's online ticket-booking behavior.

We mainly take advantage of the data at the level of the departure city, arrival city, reservation date, departure date, and travel mode, which are aggregated from the original ticket-order data from the institution. The original ticket-order data include all train and airline ticket bookings from 2017 to 2019, with detailed information on departure and arrival station, departure and arrival city, booking date, departure date, flight (train) number, price, booking platform source, and order status.[9] Our sample records 0.25 billion passengers who ever booked train tickets in 2017–2019, with 1.92 billion train trips in total, and 0.12 billion passengers for airline tickets, with 0.65 billion airline trips. In our following analysis, we focus on the completed orders, which are paid and issued finally, to capture the actual travels. We also take advantage of the orders that were canceled to explore possible behavioral channels.[10]

An order may include multiple trips and passengers—simply aggregating each trip may cause measurement problems when passengers order connected-trip or round-trip tickets; for example, an individual living in Beijing wants to book a round-trip ticket to Shanghai leaving in two days and will stay in Shanghai for three days. Simply aggregating these two trips will identify one individual leaving from Beijing to Shanghai who books a ticket for departure in two days and another one leaving from Shanghai to Beijing who books a ticket for departure in five days. Therefore, to identify independent trips, we need to deal with the round trips and connected trips. First, we identify the round-trip cases by checking whether an individual has two trips booked on the same day and for which the destination city of the first trip (ranked by the departure time) is the same as the departure city of the second trip. We delete the second trip of the round-trip tickets in our analysis because it is not an independent travel behavior. Second, we identify the connected-trip cases by checking whether two trips booked by an individual on the same day are connected by a transfer city. We combine these two trips and construct a new trip with the first trip's departure city as the original departure city and the second trip's destination city as the final-destination city. The price of this new trip is the sum of the former two trips' prices. Following the above method, we successfully identify 18% as connected trips and 10% as round trips for trips by train; for trips by air, 10% are connected trips and 24% are round trips.[11]

We group the completed orders by booking period and age cohort. The booking period is defined as the difference between the departure date and booking date, and the orders are separated into four groups: 0–1 days, 2–7 days, 8–30 days, and more than 30 days. The age cohort is divided into 0–18, 19–35, 36–55, and 56–100 years old. In total, we have 1.3 billion train tickets and 0.36 billion airline tickets in our analysis. Price is the average price of all trips for each city-pair-reservation-date cell.

We construct *NumRatio*, the propensity of daily average individuals leaving from the origin city to a given destination city on a given date in a specific booking period (leaving in 0–1, 2–7, 8–30, more than 31 days), which is calculated as the ratio of the number of ticket bookings between a city pair on a given date in a specific booking period and the average number of daily trips (in 10,000) from the origin city that year.

2.3. Air pollution

The air-pollution data are from the China National Environmental Monitoring Center (CNEMC).[12] The data consist of AQI and multiple pollutants including $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and ozone for more

[4] Source: The 44th China Statistical Report on Internet Development from China Internet Network Information Center (CNNIC). https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201908/P020190830356787490958.pdf.

[5] According to the 2018 Statistical Bulletin on the Development of the Transportation Industry of China (https://xxgk.mot.gov.cn/jigou/zhghs/201904/t20190412_3186720.html), in 2018 (2019), civil aviation transported 537 (575) million passengers on domestic routes, and railway transported 3.375 (3.66) billion passengers throughout the year—546.8 (521.4) million people booked train tickets and 122.7 (133.5) million booked domestic airline tickets in our sample of the ticket bookings provider.

[6] The website is http://www.12306.cn. 12306 China railway is an information service website affiliated with China National Railway Group Co., Ltd.

[7] https://www.qianzhan.com/analyst/detail/220/180627-5f0bf9d9.html.

[8] This information comes from a report by the Qianzhan Research Institute, https://www.sohu.com/a/399744531_99922905.

[9] Order status reflects the final status of an order and is identified by two dimensions: one is whether the ticket is issued, canceled, or changed; the other is whether the ticket is paid, refunded totally, or refunded partially.

[10] The booking date is updated for the tickets that are changed as the latest booking day. So, we do not know the original booking date for these orders and thus cannot include these tickets in our analysis.

[11] Both connected trip and round trip are identified separately in train-ticket-booking dataset and airline-ticket-booking dataset. For the cases in which an individual books a train ticket to leave and an airline ticket to return, we regard these two tickets as independent trips instead of a round trip.

[12] We downloaded the data from https://quotsoft.net/air/, which are retrieved from CNEMC.

than 1,500 stations. Because AQI and $PM_{2.5}$ are better known than the rest of the pollutants and people may be more aware of and more responsive to these two, we use AQI in our main analysis and $PM_{2.5}$ for our robustness check. We aggregate AQI and $PM_{2.5}$ into the daily level by taking the average of the hourly readings.

2.4. Weather variables

The weather data are obtained from the China Meteorological Data Service Center (CMDSC),[13] affiliated with the National Meteorological Information Center of China. The data provide daily weather conditions such as pressure, temperature, relative humidity, wind speed, wind direction, sunshine duration, and precipitation, from 745 basic and reference surface meteorological observation stations.[14]

Because some cities may have no stations, we aggregate the station-level pollution and weather data into the city level using the following procedure. If at least one station is within the city boundary, we use the mean of the station readings; if no station is in the city, we use readings of the station nearest to the city's centroid.[15]

2.5. Summary statistics

The summary statistics are shown in Table 1. We keep all ticket booking records between prefecture-level cities and 4 municipality cities and aggregate them into the city-pair-date level. We do not include the city-pair-date cells with zero booking records, under the assumption that people cannot book tickets between the specific city pair on that day if not even one ticket is booked. Hence, our sample is an unbalanced panel.

Our sample contains 321 prefecture-level cities and four municipality cities,[16] 89,451 city pairs in total, among which train tickets involve 83,644 city pairs and airline tickets involve 29,908 city pairs.[17] A city pair per day has about 33.4 and 20.2 ticket bookings departing within one day for train and airline, respectively, and 16.8 and 21.8 ticket bookings departing within two to seven days for train and airline, respectively. The AQI difference ranges from about −489 to 485, with an average of around zero (distribution of the AQI difference is shown in Appendix Figure A.1). More train ticket bookings depart in 0–1 days than in 2–7 days; by contrast, more airline tickets are booked to depart in 2–7 days. On average, a visitor from a given origin city on a given date will travel to a given destination city by train (air) within one day with a probability of 1.49% (3.73%).

3. Empirical strategy

We study the impact of the AQI gap on decisions regarding short-term trips at the city-pair-date level with the following econometric specification[18]:

$$NumRatio_{ijtr} = \beta_0 + \beta_1 (P_{it} - P_{jt}) + f(W_{it} - W_{jt}) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijtr}, \quad (1)$$

where $NumRatio_{ijtr} = \frac{Num_{ijtr}}{\sum_{t=1}^{365} Num_{it}/365}$ (the unit of the denominator is 10,000 trips); r denotes four reservation windows, 0–1, 2–7, 8–30 days, and more than 30 days, respectively; i denotes origin city; j denotes destination city; t denotes reservation date. The dependent variable is the *probability* of individuals from an origin city i choosing to travel to destination city j on reservation date t in the reservation window r. It is defined as the ratio between the number of ticket bookings leaving from city i to city j in the booking period r on a given reservation date t ($NumRatio_{ijtr}$) and the average number of trips from the origin city i per day that year ($\sum_{t=1}^{365} Num_{it}/365$). A $NumRatio_{ijtr}$ of 100 means that, on average, for every 10,000 trips from city i booked on date t, 100 of them are booked to city j in the reservation window r, or that a visitor from city i has a 1% probability of booking a ticket to city j on date t in the reservation window r. The key explanatory variable is $P_{it} - P_{jt}$, i.e. the pollution difference between city i and city j on date t. β_1, the parameter we are interested in, captures how the pollution gap between a city pair i-j affects the probability of traveling from city i to city j in the reservation window r on date t.

We further control for high-dimensional fixed effects and various weather conditions. γ_{im} and θ_{jm} are city-year-month fixed effects for departure and destination cities, respectively, to control for specific time-varying characteristics of origin cities and destination cities such as the attractiveness of the city; δ_{ij} are city-pair fixed effects that absorb the specific travel pattern between a city pair; ϕ_t are date fixed effects, to account for seasonal and specific date patterns of travels; $f(W_{it} - W_{jt})$ denotes linear and quadratic forms of weather conditions including temperature, air pressure, wind speed, and precipitation. All standard errors are two-way clustered at the origin- and destination-city level.

3.1. Instrumental variable

We use an instrumental variable approach to estimate the causal effect of pollution gaps because the pollution difference of a city-pair is endogenous for the following two reasons:

Omitted Variable. Pollution is correlated with local economic conditions and a city's natural (unobservable) attractiveness. High-pollution cities are likely to be developed cities that naturally attract visitors and business trips. We control for the time-invariant characteristics are controlled by destination-city fixed effects. However, a city's attractiveness may also vary with time; for example, a big event such as the Olympic Games would both increase the visiting attractiveness and pollution, which leads to an upper bias estimation of ordinary least squares (OLS). To rule out these confounding factors, we control for date fixed effects and city-year-month fixed effects for both departure cities and destination cities. Although we rule out most time-varying

[13] http://data.cma.cn/data/detail/dataCode/SURF_CLI_CHN_MUL_DAY.html.

[14] The unit of daily wind speed is *m/s*. The daily wind direction refers to the direction where today's maximum wind comes from, and it is by 360°; for example, a wind direction of 90° means the maximum wind today is east wind.

[15] Because we only keep prefecture-level regions and municipality cities in our sample, among which only three cities do not have pollution-monitoring stations and 42 cities do not have meteorological observation stations located within the city boundary, we believe how the cities are dealt with will have little effect on the results much. We also perform analysis by excluding these cities without monitoring stations within their boundaries and find the results are very similar to our main results.

[16] Prefecture-level regions include prefectures, autonomous prefectures, prefecture-level cities, and leagues. For simplicity, we refer to "prefecture-level region" as "city" in our paper.

[17] Our instrumental variable is built at the city-pair-date level. By construction, some origin (destination) cities do not have upwind nearby cities after we exclude the counterpart destination (origin) cities, and thus do not have instrumental variables. We remove these city pairs from our sample to make the OLS and IV results comparable, which would not hurt much, reducing the number of cities from 331 to 325 and city pairs from 91,833 to 89,451.

[18] We assume the choice of ticket booking platforms is unaffected by the difference in pollution between the origin and destination. In another word, the platform choice of a trip between a city pair is independent of pollution differences. To implicitly test this assumption, we interact AQI difference with the week of Double Eleven, the largest online shopping carnival in China, during which platforms offer attractive promotions. We instrument the interaction term using the interaction of upwind pollution difference and the Double-Eleven week, and then conduct 2SLS estimation. If the platform choice is correlated with air pollution difference, we would thus expect a significant coefficient of the interaction term of air pollution difference and Double Eleven. As shown in Table A.1, the coefficients of the interaction are economically and statistically insignificant, which could support our platform independence assumption.

Table 1
Summary statistics (sample period: 2017–2019).

Variable	0–1 days					2–7 days				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Panel A: train										
# of ticket bookings	22,793,850	33.39	165.20	1	12,328	19,027,114	16.80	63.09	1	5,525
Ratio of ticket bookings	22,793,850	149.26	610.61	0.30	186,958.23	19,027,114	179.04	693.78	0.52	359,753.59
Panel B: flight										
# of ticket bookings	5,471,235	20.20	61.676	1	2,820	6,679,679	21.78	58.10	1	2,166
Ratio of ticket bookings	5,471,235	373.20	1,471.40	1.01	1,095,000	6,679,679	305.66	1224.45	0.85	1,042,857.10
Panel C: pollution and weather difference (all)										
AQI difference	25,055,837	−0.14	47.22	−481.44	484.45	22,274,124	0.11	48.37	−488.96	483.42
AQI of origin city	25,056,088	69.57	43.11	1	500	22,274,380	68.82	43.23	1	500
AQI of destination city	25,056,114	69.71	42.99	1	500	22,274,373	68.71	43.29	1	500
Temperature diff (°C)	25,323,932	−0.01	6.27	−53.55	54.15	22,520,218	−0.02	6.79	−55.15	55.45
Precipitation diff (mm)	25,351,376	0.01	12.59	−1529.90	1529.90	22,544,257	0.003	13.03	−1529.90	1529.90
Wind speed diff (m/s)	25,320,542	0.002	1.36	−12.95	12.95	22,517,446	−0.003	1.38	−12.85	12.95
Atmospheric pressure diff (hPa)	25,322,281	−0.35	72.91	−434.32	433.23	22,518,578	0.42	78.28	−434.32	433.23
Sunshine diff (h)	25,159,874	0.01	4.62	−19	19	22,384,227	−0.01	4.72	−19	19
Relative humidity diff (%)	25,324,790	−0.01	19.67	−92.45	92.45	22,520,647	0.07	20.24	−92.45	90
Upwind AQI difference (IV)	25,360,466	−0.15	11.99	−224.95	215.85	22,552,181	−0.01	12.21	−224.95	215.85

Notes. All the statistics are reported for each city pair per day during the sample period 2017–2019. Panels A and B report the summary statistics for train and airline ticket bookings, respectively. *# of ticket bookings* is the number of ticket bookings between a city pair per day. *Ratio of ticket bookings* is the ratio of the number of individuals traveling from a given origin city to a given destination city on a given date in a specific booking period and the average daily number of trips (in 10,000) traveling from the origin city of that year. *AQI difference* is the difference between AQI in the origin city and that in the destination city. *Upwind AQI difference* is the difference between upwind AQI (constructed using cities nearby) of the origin city and that of the destination city.

attractiveness by controlling for city-month fixed effects, we cannot completely mitigate the potential short-term time-varying attractiveness of a city.

Reverse Causality. Pollution is correlated with economic activities—not only do trains or airplanes emit pollutants during travels, but visitors also “bring” pollution out of the departure city and into the destination city through their activities such as taking a taxi or public transportation, and thus narrow the pollution gap between the two cities. This narrowing of the gap will lead to an underestimation of the pollution effect if not considered. This reverse causality problem is less important because we focus on the ticket *reservation* date and the reservation itself would not bring pollution to the destination city. For the ticket bookings that depart within one day, however, this problem may still cause the underestimation of the effects.

We take advantage of the spatial spillover characteristic of air pollution to construct the instrument variable for the local air pollution (Jia and Ku, 2019; Zhang et al., 2015). Spatial spillover of air pollution is widely found in the literature, for example, Zhang et al. (2015) find about half of Beijing’s pollution is from sources outside of the city municipality. Additionally, wind speed is a major facilitator in transporting $PM_{2.5}$ (Wang et al., 2017). Also, daily travel bookings between two cities are not correlated with the wind that day in their respective neighboring cities, which satisfies the exclusion restriction on the IV. In this way, the pollution of a city can be divided into two parts: one part is the pollution generated by its own production and living activities, and the other part is the pollution from other neighboring cities blown by the wind. Following Chen et al. (2021) and Barwick et al. (2024), we utilize the second part to construct the IV, based on the exogenous daily variation of wind direction and wind speed. Let us take city i as an example, and the same goes for city j. Upwind pollution in city i at reservation date t (UP_{it}) is the sum of pollution carried by wind in nearby upwind cities n (excluding counterpart city j) within a given radius from city i, adjusted by wind speed (WS_{nt}) and the inverse squared distance between the city i and city n ($1/D_{ni}^2$):

$$UP_{it} = \sum_{n \neq i} \frac{WS_{nt} \times cos(\lambda_{ni} - \rho_{nt})}{D_{ni}^2} \times 1[\textit{city n lies upwind of city i}]_{nt} \times P_{nt}, \tag{2}$$

where city n is the nearby city within a given radius of city i, P_{nt} is the pollution of city n on date t, WS_{nt} is the wind speed in city n on date t, and

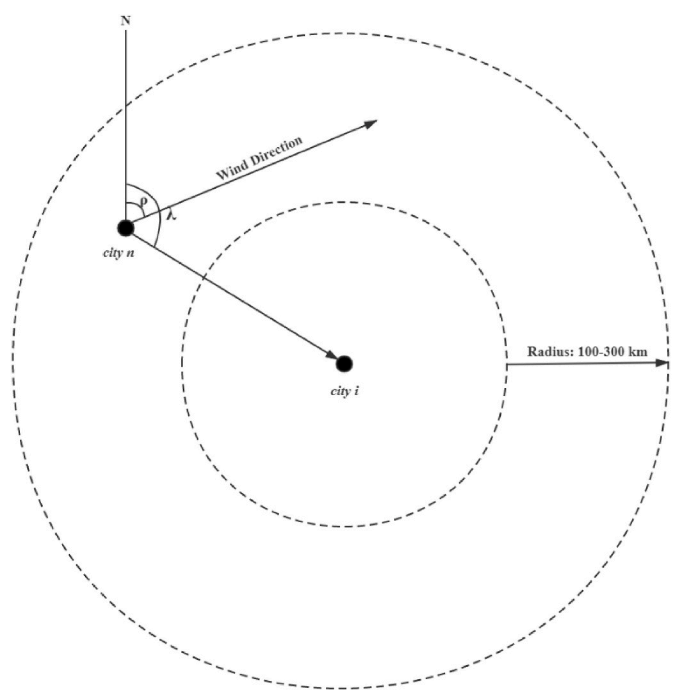


Fig. 1. IV construction
Notes. This figure illustrates how the IV is constructed. City n is the nearby city of city i within a specific radius. ρ is the wind direction of city n, which has a daily variation; λ is the angle between the north direction and the straight line between the nearby city n and city i.

D_{ni}^2 is the squared distance between city n and city i. As illustrated in Fig. 1, λ_{ni} is the angle between the north direction and the straight line between nearby city n and city i; ρ_{nt} is the wind direction (ranging from 0 to 2 π) of city n on date t. $1[\textit{city n lies upwind of city i}]_{nt}$ is a dummy equals to one if the nearby city n lies upwind of city i on date t, otherwise zero. A negative $cos(\lambda_{ni} - \rho_{nt})$ means the city n is located downwind of city i; then, pollution in city n will not be a source of pollution in city i.

We utilize the non-local variation of pollution in city i, which is the part carried by the wind from upwind cities within a specific radius, and then adjust it using wind speed in city n on date t and the inverse squared distance. Upwind cities of city i are chosen based on the daily wind direction of all nearby cities of city i within a given radius—wind direction varies daily, and thus the list of upwind cities for a given city also changes every day. Overall, the variation of our IV comes from the wind speed of upwind city n, the wind direction of upwind city n, and the distance between city i and upwind city n, which all have daily variation.

If city i and nearby city n are too close, city n may affect city i through channels other than just pollution transmission; therefore, to make the exclusion restriction hold, we cannot make the radius too small. We also cannot make the radius too large, because the prediction power would be insufficient when the upwind city n is too far from city i. Therefore, following Chen et al. (2021), we mainly use cities within the 100 km–300km buffer-zone radius of city i as the candidates for upwind cities and use other radii for robustness checks (coefficients reported in Table A.4). To further avoid the exclusion restriction being violated, we exclude the destination city j from the candidate list if city j is one of the nearby upwind cities of city i on date t, and exclude origin city i from the candidate list if city i is one of the nearby upwind cities of city j on date t.

Figure A.2 Panel A presents the correlation between upwind AQI and local AQI at the city level. Then, we use the upwind pollution difference ($UP_{it} - UP_{jt}$) as the IV of the pollution gap ($P_{it} - P_{jt}$). Figure A.2 Panel B shows the raw correlation between upwind AQI difference and AQI difference at the city-pair level, without controlling any fixed effects. We then perform two-stage least-squares (2SLS) estimation:

$$1^{\text{st}}\ \text{stage}: P_{it} - P_{jt} = \delta_0 + \delta_1 \left(UP_{it} - UP_{jt}\right) + f\left(W_{it} - W_{jt}\right) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \epsilon_{ijt}, \tag{3}$$

$$2^{\text{nd}}\ \text{stage}: NumRatio_{ijtr} = \beta_0 + \beta_1 \left(\widehat{P_{it} - P_{jt}}\right) + f\left(W_{it} - W_{jt}\right) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijtr}. \tag{4}$$

4. Results

4.1. Baseline results

Table 2 shows our baseline results for all travel bookings in columns (1)–(2), train ticket bookings in columns (3)–(4), and airline bookings in columns (5)–(6). We control for weather conditions, city-pair fixed effects, date fixed effects, origin-city-year-month fixed effects, and destination-city-year-month fixed effects in all columns. The standard errors are two-way clustered at the origin- and destination-city level. The relative effect is reported as 50 times the estimated coefficient's value divided by the average of the dependent variable in each column, i.e. the percentage change caused by a 50-unit increase in AQI difference.

Panel A of Table 2 reports the results from 2SLS estimation in equations (3) and (4), and Panel B shows the OLS estimates in equation (1). The KP-F statistics are all more than 145 for all groups, which reject the existence of the weak-IV problem. As expected, the 2SLS estimates are larger than the OLS estimates because of possible reverse-causality issues which would bias the OLS estimates toward zero, especially for the trips departing within a day.

We find that a 50-unit increase in the AQI gap (the standard deviation of the AQI gap ranges from 45 to 60 units across different groups) between an origin and a destination city significantly increases ticket bookings (train and airline) from the origin city to the destination city departing within one day and in two to seven days by 1.30% and 1.33%, respectively.[19] Additionally, the effects are slightly larger on train tickets than on airline tickets. Regarding different travel modes, a 50-unit increase in the AQI gap between a city pair leads to a 1.26% and 1.13% increase in train ticket bookings departing within a day and two to seven days, whereas the corresponding effects are 1.12% and 0.83% for flight bookings. The standard errors in the above analyses are two-way clustered at the origin- and destination-city level and the results are robust when one-way clustering the standard errors at the city pair level, with results reported in Appendix Table A.3.

Next, we examine the nonlinear relationship between the AQI difference and ticket bookings. Several papers have found the nonlinear effects of air pollution and individuals' response to severe air pollution is disproportionally higher than their response to high pollution (Chen et al., 2021; Chen et al., 2018; Qin et al., 2019; Qin and Zhu, 2018). Therefore, we follow the literature and decompose the continuous AQI gap into six bins (AQI difference <35, in [35,50), [50,75), [75,150), [150,250), and ≥ 250, respectively). As shown in Fig. 2, the effect gets disproportionally larger as the AQI gap widens: the magnitude of the coefficients increases with the pollution gap, indicating the nonlinear effect of the air-pollution difference between cities on travels between cities. Specifically, from the coefficients reported in Appendix Table A.2, we show that compared with the base group (AQI difference less than 35), ticket bookings between a city pair will increase by 9.6%, and 4.5% departing within one day and in two to seven days, respectively, if the AQI in origin city is at least 250 units higher than the AQI in the destination city.

We also conduct several robustness checks. First, as a falsification test, we show in Table 3 that the AQI difference between a city pair does not affect ticket bookings departing in more than one week, which lends more support to our research design that our finding is not driven by some mechanical correlation between the city-pair-date-level unobservables and travel bookings.

Second, we replace AQI with $PM_{2.5}$ as the measure of air quality and find similar results for overall and train ticket bookings. As shown in Table 4, the effects of the $PM_{2.5}$ difference are slightly larger for airline ticket bookings than the effects of AQI: a 50-unit increase of $PM_{2.5}$ gap between a city pair leads to a 1.59% and 1.02% increase in airline ticket bookings departing in within one day and in two to seven days, respectively.

Third, to ensure our results are robust to how we deal with zero cells, we further keep all the city-pair-date cells and fill those with no records of ticket bookings with zero, with the assumption that everyone can book tickets between any city pair at any date. The coefficients are shown in Table 5 and their values are similar to the baseline estimates.

Additionally, to investigate whether origin and destination pollution have different effects, we include air pollution in the origin and destination separately on the right-hand side of the equation as a horserace model. As shown in Table 6, a 50-unit increase in AQI in origin cities will increase the train ticket bookings departing within one day (2–7 days) by 2.01% (1.75%), whereas the destination pollution does not have a statistically significant effect, which is in line with our expectation that

[19] 1.30% = 0.036*50/139.582, 1.33% = 0.042*50/157.025. The following calculation of the relative effect of a 50-unit increase in AQI follows the same logic. We also use the average number of daily ticket *searches* leaving from city i that year as the denominator of the dependent variable, instead of the average number of daily ticket bookings, to test whether the choice of denominator may change our results. Similarly, we find a 50-unit increase in AQI leads to a 2% increase in the ticket bookings leaving within one day, whereas it has no effect on those leaving in more than one day.

Table 2
Main results.

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
	Dependent variable: AQI difference					
Panel A: 2SLS						
1st stage						
Upwind AQI difference	1.687***	1.663***	1.671***	1.634***	1.684***	1.695***
	(0.110)	(0.112)	(0.108)	(0.108)	(0.139)	(0.137)
KP F stat.	233.6	220.6	240	228.2	146.9	152.8
	Dependent variable: Ratio of ticket bookings					
2nd stage						
AQI difference	0.036***	0.042***	0.038***	0.041**	0.083*	0.050***
	(0.006)	(0.013)	(0.008)	(0.017)	(0.044)	(0.018)
Benchmark	139.582	157.025	149.355	178.998	372.117	304.989
Relative effect (%)	1.301	1.334	1.263	1.134	1.117	0.828
	Dependent variable: Ratio of ticket bookings					
Panel B: OLS						
AQI difference	0.002	0.011***	0.004	0.016**	0.017	0.007**
	(0.002)	(0.004)	(0.003)	(0.007)	(0.021)	(0.003)
Benchmark	139.582	157.025	149.355	178.998	372.117	304.989
Relative effect (%)	0.076	0.359	0.124	0.451	0.232	0.117
Observations	24,806,951	22,063,482	22,292,946	18,608,699	5,355,666	6,541,710
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the impact of air pollution on travel decisions. Panel A reports the 2SLS estimates of equations (3) and (4), and Panel B reports the OLS estimates of equation (1). For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather differences between the origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

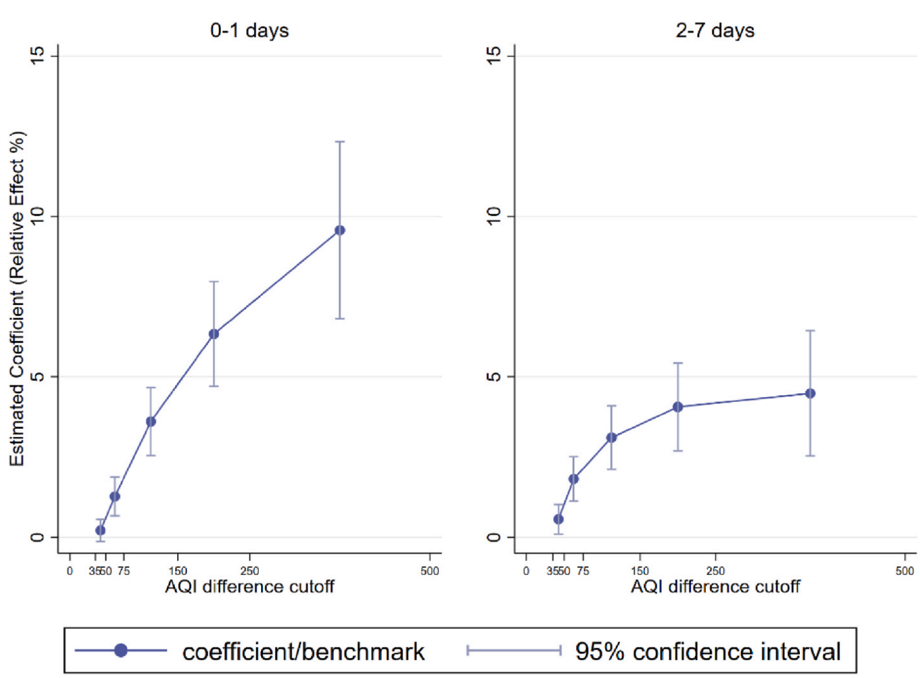


Fig. 2. Non-linearity
Notes. This figure plots the nonlinear effects of the AQI difference on ticket bookings. The dependent variable is the propensity of the daily average individuals leaving from the origin city to a given destination city on a given date in a specific booking period. Each dot denotes the relative effect (%) calculated by dividing the estimated coefficient (reported in Table A.2) of each bin of the AQI difference and the benchmark value of the dependent variable. The bars are the 95% confidence intervals.

Table 3
Falsification test–ticket bookings in more than 7 Days.

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	8–30 days	>30 days	8–30 days	>30 days	8–30 days	>30 days
	Dependent variable: AQI difference					
Panel A: 2SLS						
1st stage						
Upwind AQI difference	1.684***	1.720***	1.662***	1.685***	1.706***	1.743***
	(0.117)	(0.120)	(0.113)	(0.101)	(0.136)	(0.134)
KP F stat.	207.7	204.8	217.2	278.7	157.1	169.6
	Dependent variable: Ratio of ticket bookings					
2nd stage						
AQI difference	−0.008	0.163	−0.005	0.331	0.020	0.133
	(0.024)	(0.219)	(0.037)	(0.606)	(0.021)	(0.098)
Benchmark	211.080	893.830	247.876	1849.104	367.998	789.255
Relative effect (%)	−0.179	0.912	−0.111	0.896	0.266	0.842
	Dependent variable: Ratio of ticket bookings					
Panel B: OLS						
AQI difference	−0.027*	−0.070	−0.037	−0.060	0.0005	−0.052
	(0.016)	(0.135)	(0.022)	(0.314)	(0.010)	(0.056)
Benchmark	211.080	893.830	247.876	1849.104	367.998	789.255
Relative effect (%)	−0.642	−0.394	−0.737	−0.162	0.006	−0.326
Observations	16,443,826	3,788,614	13,467,361	1,739,837	5,406,839	2,465,881
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the falsification test using tickets booked to depart in more than one week. Panel A reports the 2SLS estimates of equations (3) and (4), and Panel B reports the OLS estimates of equation (1). For each column, the dependent variable is the probability of individuals from an origin city i choosing to travel to destination city j on reservation date t in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked departing in 8–30 days, and columns (2), (4), and (6) report the results for tickets booked to depart in more than 30 days. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

higher pollution in the origin city plays a major role in pushing people to flee high-polluted cities. For flight bookings, we find slightly different results that the pulling from the destination plays a larger role than the pushing from the origin, shown in columns (5)–(6).

Furthermore, to ensure the robustness of our findings concerning the choice of the dependent variable, we conduct additional robustness checks using alternative definitions of the dependent variable. Firstly, we replace the denominator in our original dependent variable with different measures of the origin city's daily trips, which remain constant across years. This could reserve the variance of total trips across origin city-year. We employ two alternative measurements as the denominator: 1) the average daily number of trips over the three-year period (2017–2019) for ratio2, and 2) the average daily number of trips in the first year of the sample (2017) for ratio3, to measure the *probability* of individuals from an origin city i choosing to travel to city j on date t. As shown in Appendix Table A.5, a 50-unit increase in the AQI difference between a city pair leads to a 1.313% (1.380%) increase in trips leaving within 0–1 (2–7) days when using ratio2, and a 1.427% (1.197%) increase in trips leaving within 0–1 (2–7) days when utilizing ratio3. The magnitudes of these coefficients are notably consistent with our baseline results. In addition, we replace the dependent variable with the logarithm of the number of trips, as another robustness check. The results, presented in Appendix Table A.6, are similar to our baseline findings, further validating the robustness of our results.

4.2. Heterogeneity

The detailed travel information, especially the traveler's personal information, allows us to investigate the heterogeneities of the main results. If the increased trips are indeed induced by pollution, then these trips are more likely to be relatively low-cost travel for leisure purposes. We would expect the effects of pollution to be greatest for groups with lower short-term travel costs, and for trips traveling to destinations that are geographically closer—allowing for travel completion within two days—and to destinations that are, on average, cleaner.

First, we are interested in how people of different ages respond to pollution differently. Williams (2019) also indicates that the decisions to engage in defensive behaviors are significantly correlated with age. Fig. 3 shows the estimated relative effects for different age groups by dividing our sample into four age groups: 0–18 years old, 19–35 years old, 36–55 years old, and more than 55 years old. The coefficients are reported in Appendix Table A.7.[20] We find the effect is the strongest for people ages 19–35 and declines in older cohorts. For different age groups, we need to consider not only people's awareness of air-pollution hazards but also the willingness and feasibility to avoid pollution by taking short-term trips in terms of pecuniary cost and time cost. Although children and old people are the most vulnerable groups, they may bear higher costs for short-term trips than middle-aged people: the old may find it more troublesome to make short-term trips to avoid pollution compared with staying indoors or wearing masks, and children have less flexibility to travel during semesters. Additionally, middle-aged people may prefer outdoor activities and thus value outdoor air quality more than old people do.

We further investigate where those people are going, by dividing our sample into intra-province and inter-province trips, which could support our hypothesis that people may want *short-term* trips to avoid pollution. As shown in Fig. 4 (with coefficients reported in Appendix Table A.8),

[20] The benchmark value we used to compute the relative effect is the *average propensity* of the *specific age group* from city i at time t traveling to city j; for example, 639 (Table A.7, col. 1) means that, on average, people ages 0–18 years old outflowing from origin city i will have a 6.39% probability of traveling to destination city j on date t.

Table 4
Robustness – $PM_{2.5}$ as the measure of air pollution.

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
	Dependent variable: $PM_{2.5}$ difference					
Panel A: 2SLS						
1st stage						
Upwind $PM_{2.5}$ difference	2.234***	2.209***	2.212***	2.168***	2.252***	2.271***
	(0.126)	(0.130)	(0.123)	(0.126)	(0.166)	(0.164)
KP F stat.	312.4	288.3	321.2	297.4	184.7	191.3
	Dependent variable: Ratio of ticket bookings					
2nd stage						
$PM_{2.5}$ difference	0.037***	0.043***	0.036***	0.037*	0.118*	0.062***
	(0.007)	(0.015)	(0.009)	(0.020)	(0.063)	(0.019)
Benchmark	139.579	157.022	149.354	178.996	372.111	304.984
Relative effect (%)	1.341	1.359	1.210	1.035	1.592	1.015
	Dependent variable: Ratio of ticket bookings					
Panel B: OLS						
$PM_{2.5}$ difference	0.004	0.015**	0.007	0.020**	0.022	0.008**
	(0.003)	(0.006)	(0.006)	(0.009)	(0.029)	(0.003)
Benchmark	139.579	157.022	149.354	178.996	372.111	304.984
Relative effect (%)	0.147	0.477	0.235	0.562	0.292	0.137
Observations	24,806,706	22,063,213	22,292,807	18,608,569	5,355,522	6,541,532
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the robustness check of using $PM_{2.5}$ as an alternative measure of air pollution. Panel A reports the 2SLS estimates of equations (3) and (4), and Panel B reports the OLS estimates of equation (1). For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions for control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

we find the effects are driven by people who book train tickets for intra-province trips and depart within one day, but not for airline tickets. The difference in the estimated effects for intra-province and inter-province trips are statistically significant, based on Fisher's permutation test. Next, we further divide the train ticket bookings into three groups according to the number of 4 A-plus tourist attractions[21] of the destination city. Fig. 5 (with coefficients reported in Appendix Figure A.9) shows that people are departing to cleaner intra-province cities with more tourist attractions within one day and to cleaner cities with more tourist attractions outside the province in two to seven days, in line with the purpose of pollution avoidance. In our sample, the median length of a round-trip by train is about zero to one days, which is consistent with our findings of *short-term* avoidance of air pollution exposure—people tend to travel more by train than by air to an intra-province city with more tourist attractions when their origin city is more polluted than the destination city.

In addition, we examine the heterogenous responses of pollution-induced trips across different average pollution levels of origin or destination cities. Specifically, for every city in our sample, we count the number of days between 2017 and 2019 where the daily average AQI surpassed the threshold 100, which is commonly used as a threshold for a polluted city. We then divide the sample into two categories based on the median number of days exceeding this threshold, resulting in two groups each comprising 173 cities. As shown in Table 7, the effects of AQI difference between a city pair are more prominent if the origin city is on average a high-pollution city, and if the destination city is on average a clean city. Specifically, if the origin city has higher pollution levels, a 50-unit increase in the AQI difference leads to a 1.452% (1.669%) increase in trips between the city pair departing within 0–1 (2–7) days. In contrast, for origin cities with lower pollution levels, this effect decreases to 0.811% and 0.681% respectively for the same departure intervals. Moreover, for cleaner destination cities, a 50-unit increase in the AQI difference leads to a 2.209% (1.946%) increase in trips between the city pair departing within 0–1 (2–7) days. Conversely, for less clean destinations, the effects diminish to 1.117% and 1.224% respectively for the same time frames. It is important to note that a "more polluted city" refers to a city that, *on average*, experiences higher pollution levels compared to other cities over the sample period, but this does not necessarily mean that the city is more polluted on any given day.

5. Mechanisms and further discussions

5.1. Rational and behavioral factors in trip decision-making process

Two types of decision-making factors could be behind the pollution-induced trips: rational and behavioral considerations. Travel decisions could be rational, which means that the trips are planned to avoid pollution exposure, based on air pollution forecasts for the intended date of departure. These forecasts may originate from either official forecasts released to the public by governmental agencies, or travelers' personal predictions, informed by current pollution levels and the observed tendency for air pollution to persist over time. Alternatively, decisions may be induced by emotional shocks: people in bad environments are likely to book tickets to reduce stress because air pollution reduces happiness (Zhang et al., 2017); or people are just more irrational and sensitive to

[21] According to the Standard of Rating for Quality of Tourist Attractions (GB/T 17775-2003) of the People's Republic of China, all tourist attractions can be divided into five levels, from highest to lowest as AAAAA, AAAA, AAA, AA, A-level tourist attractions. Our data contain 261 AAAAA tourist attractions and 3,659 AAAA ones—we denote AAAA and AAAAA as "4 A-plus" tourist attractions.

Table 5
Robustness – fill zero for lack-of-observed booking records.

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
	Dependent variable: AQI difference					
Panel A: 2SLS						
1st stage						
Upwind AQI difference	1.791***	1.791***	1.779***	1.779***	1.712***	1.712***
	(0.110)	(0.110)	(0.107)	(0.107)	(0.138)	(0.138)
KP F stat.	266.9	266.9	274.2	274.2	154.7	154.7
	Dependent variable: Ratio of ticket bookings					
2nd stage						
AQI difference	0.026***	0.030***	0.024***	0.028**	0.067	0.044**
	(0.005)	(0.010)	(0.006)	(0.011)	(0.043)	(0.022)
Benchmark	101.148	101.215	108.909	108.986	238.832	238.923
Relative effect (%)	1.265	1.504	1.108	1.279	1.398	0.930
	Dependent variable: Ratio of ticket bookings					
Panel B: OLS						
AQI difference	0.002	0.007**	0.003	0.009**	0.008	0.002
	(0.002)	(0.003)	(0.002)	(0.004)	(0.016)	(0.004)
Benchmark	101.148	101.215	108.909	108.986	238.832	238.923
Relative effect (%)	0.076	0.330	0.130	0.413	0.159	0.050
Observations	34,237,113	34,237,113	30,574,889	30,574,889	8,356,476	8,356,476
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the robustness check of replacing the cells of city pair-date that have no ticket booking records with zero. Panel A reports the 2SLS estimates of equations (3) and (4), and Panel B reports the OLS estimates of equation (1). For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

travel promotions on polluted days. A few papers have documented that air pollution leads to cognitive biases in decision-making (Chang et al., 2018; Chew et al., 2021; Li et al., 2019; Qin et al., 2019; Zhang et al., 2018). Notably, if individuals exhibit projection bias (Busse et al., 2015; Conlin et al., 2007) on polluted days, they may incorrectly predict their utility of travel in the near future and regret their decisions afterward.

To examine these factors, we first investigate whether individuals experience regret after booking tickets on days with a high pollution gap. Fortunately, our data also contain the refund and cancellation information of the ticket bookings, which allows us to examine whether travelers regret the decision. The raw data contain a ticket status label indicating whether the ticket has been refunded or canceled,[22] by which we can retrieve refund information for train tickets and both refund and cancellation information for airline tickets. Our sample from 2017 to 2019 includes about 52.6 million refunds and 34.5 million cancellations of airline tickets, and 221.4 million refunds of train tickets. We take the union of the city-pair-date cells of successful reservations, refunds, and cancellations as the city-pair dates for which a refund or cancellation can take place, and then replace those city-pair dates that have no refund or cancellation records with zero. The underlying assumption here is that people can receive a refund for or cancel their orders on the day they make a successful reservation.

[22] We filtered tickets that are refunded or canceled under the following criteria: for a train ticket, if the payment status is "paid" and the ticket status is "refunded," we categorize it as a refunded ticket; for an airline ticket, if the payment status is "unpaid" and the ticket is not printed, we categorize it as a cancellation ticket, and if the ticket status is "refunded," we categorize it as a refunded ticket. In brief, refund means someone placed an order and paid for the ticket but received a refund for it later. Cancellation means someone placed an order but cancelled it before payment.

Similar to our main analysis, we study the refund and cancellation behavior under the following specification:

$$\begin{aligned}1^{\text{st}}\text{ stage} : & P_{it} - P_{jt} \\ & = \delta_0 + \delta_1 \left(UP_{it} - UP_{jt}\right) + f\left(W_{it} - W_{jt}\right) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t \\ & \quad + \epsilon_{ijt},\end{aligned} \tag{5}$$

$$\begin{aligned}2^{\text{nd}}\text{ stage} : & \mathit{RefundRatio}_{ijt} \\ & = \theta_0 + \theta_1 \left(\widehat{P_{it} - P_{jt}}\right) + f\left(W_{it} - W_{jt}\right) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijt},\end{aligned} \tag{6}$$

where $\mathit{RefundRatio}_{ijt} = \frac{\mathit{NumRefund}_{ijt}}{\mathit{Num}_{ijt} + \mathit{NumRefund}_{ijt}}$.[23]

$\mathit{NumRefund}_{ijt}$ is the number of trips who originally booked a ticket on date *t* to travel from city *i* to city *j* but received a refund later; Num_{ijt} is the number of trips who booked a ticket on date *t* to travel from city *i* to city *j* and successfully traveled. This indicator measures the *probability* of refunds and cancellation for trips that are booked on booking date *t* to travel from origin *i* to destination *j*. If the irrational factors play a role in individuals' decisions to travel to cleaner cities when exposed to high pollution, individuals may regret their decision later; thus, we would expect the coefficient of the pollution gap to be positive.

As shown in Table 8, those train tickets booked on days with a 50-unit-higher AQI gap between origin and destination cities have, on average, a 1.64 % percent higher probability of being refunded subsequently, indicating that individuals regret their decisions on high-AQI

[23] We replace the dependent variable with $\frac{\mathit{NumRefund}_{ijt}}{\mathit{Num}_{ijt}}$ and the results are robust as those using $\frac{\mathit{NumRefund}_{ijt}}{\mathit{Num}_{ijt} + \mathit{NumRefund}_{ijt}}$.

Table 6
Robustness – origin and destination pollution.

	(1)	(2)	(3)	(4)	(5)	(6)
	Ratio of ticket bookings					
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
Panel A: 2SLS						
AQI origin	0.052***	0.056***	0.060***	0.063***	0.024	0.035**
	(0.009)	(0.015)	(0.010)	(0.019)	(0.017)	(0.015)
AQI destination	−0.019**	−0.032	−0.016	−0.025	−0.129	−0.062**
	(0.009)	(0.023)	(0.012)	(0.029)	(0.086)	(0.027)
Benchmark	139.449	156.998	149.195	178.931	372.120	304.993
Relative effect of origin (%)	1.874	1.776	2.014	1.752	0.316	0.573
Relative effect of destination (%)	−0.672	−1.029	−0.520	−0.699	−1.739	−1.016
KP F stat.	119.7	111.8	123.2	115.4	52.89	74.87
Panel B: OLS						
AQI origin	0.012***	0.016***	0.016***	0.025***	−0.001	0.000005
	(0.004)	(0.005)	(0.005)	(0.007)	(0.010)	(0.006)
AQI destination	0.007	−0.010	0.007	−0.012	−0.031	−0.013**
	(0.004)	(0.009)	(0.006)	(0.013)	(0.040)	(0.006)
Benchmark	139.449	156.998	149.195	178.931	372.120	304.993
Relative effect of origin (%)	0.435	0.502	0.544	0.695	−0.010	0.00009
Relative effect of destination (%)	0.251	−0.304	0.250	−0.333	−0.423	−0.212
Observations	25,001,049	22,225,854	22,468,239	18,743,784	5,394,328	6,590,030
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents results of including both air pollution in origin and destination cities as independent variables. Panels A and B report the 2SLS and OLS estimates, respectively. For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather conditions in both origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient of origin city's air pollution and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

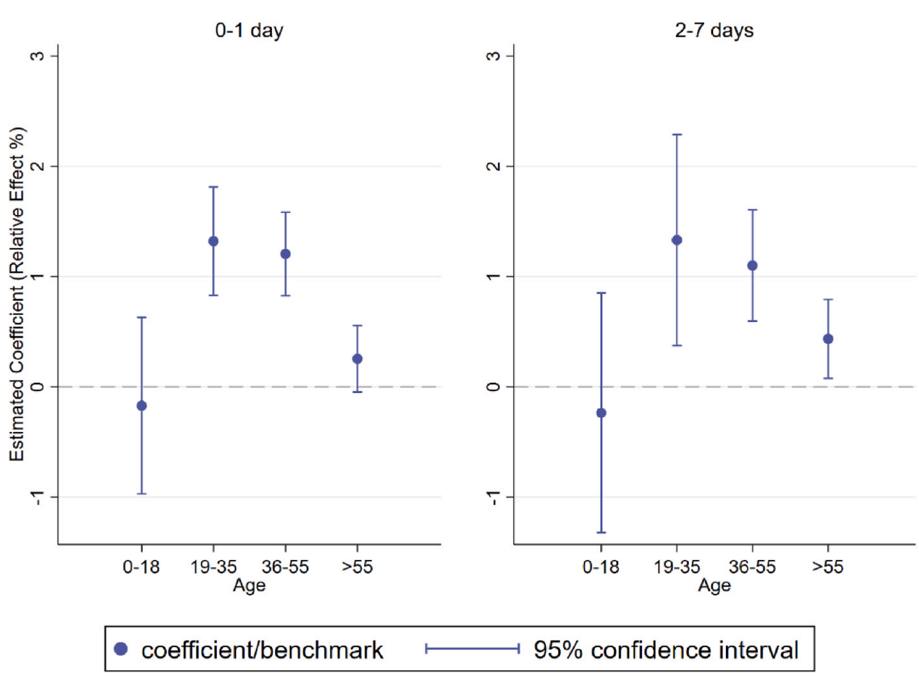


Fig. 3. Heterogeneity in Age
Notes. This figure plots the relative effects and the 95% confidence intervals for different age cohorts. The figure on the left (right) shows the results for tickets booked to depart in 0–1 days (2–7 days). Each dot denotes the relative effect (%) of a 10-unit increase in the AQI gap, which is calculated by dividing the estimated coefficient (reported in Table A.7) of the AQI difference by the benchmark value of the dependent variable and then multiplying by 50.

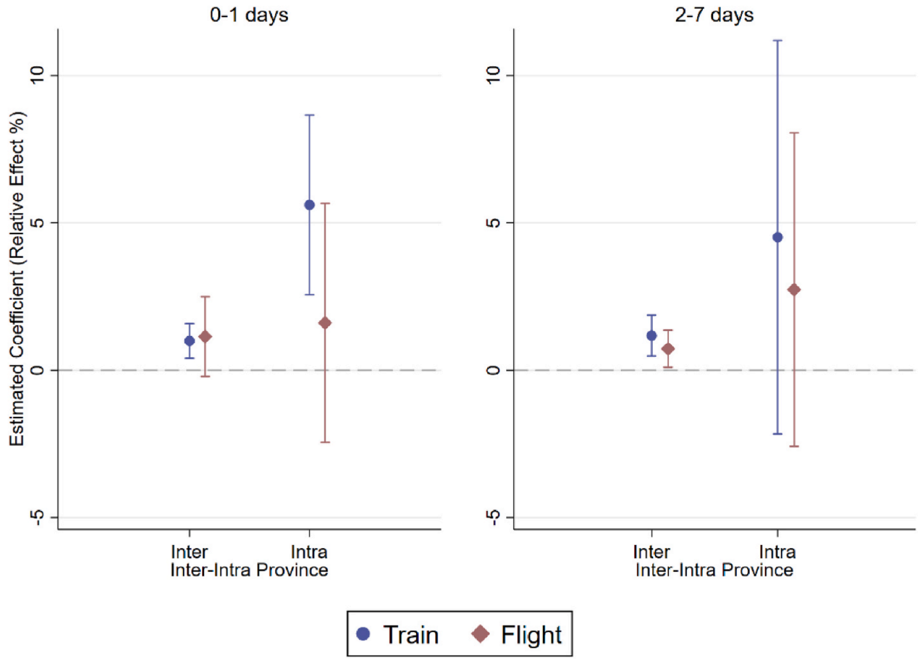


Fig. 4. Heterogeneity, Intra-province and Inter-province
Notes. This figure plots the relative effects and the 95% confidence intervals for intra-province and inter-province trips. The figure on the left (right) shows the results for tickets booked to depart in 0–1 days (2–7 days). Blue and red bars denote effects on train and airline ticket bookings, respectively. Each dot denotes the relative effect (%) of a 50-unit increase in the AQI gap, which is calculated by dividing the estimated coefficient (reported in Table A.8) of the AQI difference by the benchmark value of the dependent variable and then multiplying by 50. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

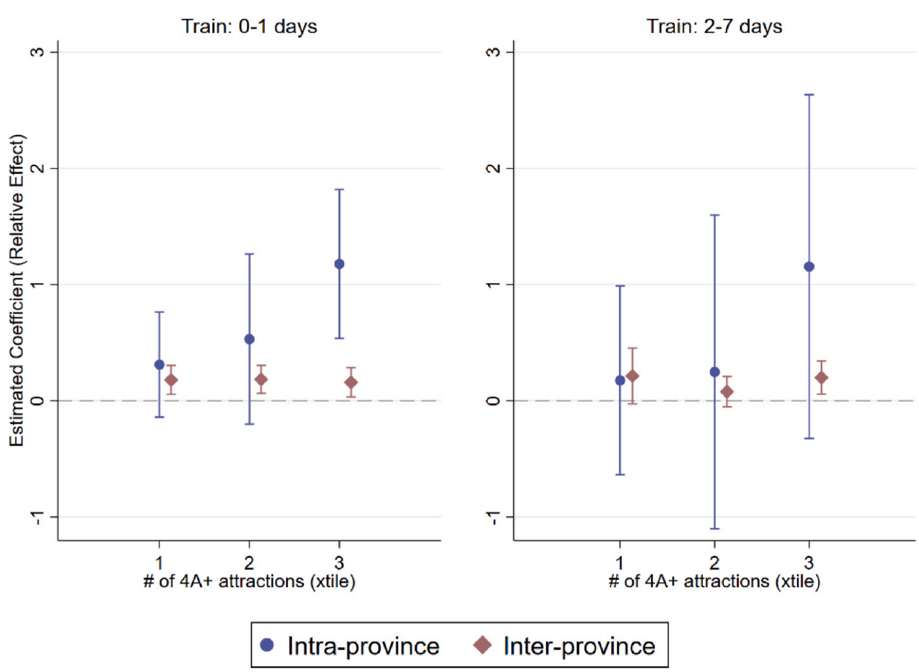


Fig. 5. Heterogeneity, 4 A-plus Tourist Attractions (Train)
Notes. This figure plots the heterogeneous relative effects and the 95% confidence intervals for train tickets, by dividing the sample into six groups according to the number of tourist attractions in the destination city and whether the trip is intra-province or inter-province. The figure on the left (right) shows the results for tickets booked to depart in 0–1 days (2–7 days). A larger xtile means more 4 A-plus tourist attractions in the destination city. Each dot denotes the relative effect (%) of a 50-unit increase in the AQI gap, which is calculated by dividing the estimated coefficient (reported in Table A.9) of the AQI difference by the benchmark value of the dependent variable and then multiplying by 50.

gap days. This increase in refund rates aligns with the theories of behavioral factors such as projection bias (Busse et al., 2015; Conlin et al., 2007). Specifically, if individuals exhibit projection bias, they may mis-predict their future utility of travel and subsequently regret their decisions. We did not find such refunding behavior in airline ticket bookings, probably because the refund fee for airline tickets is much

Table 7
Heterogeneity – in pollution level of origin or destination city.

	(1)	(2)	(3)	(4)
	0–1 days		2–7 days	
	Less Polluted	More Polluted	Less Polluted	More Polluted
Panel A: by origin's pollution level				
AQI difference	0.027**	0.035***	0.025	0.046**
	(0.014)	(0.008)	(0.016)	(0.020)
Coefficient difference		−0.008		−0.022***
p-value		0.100		0.000
Benchmark	166.268	121.207	182.115	138.931
Relative effect (%)	0.811	1.452	0.681	1.669
KP F stat.	169.5	179.3	169.7	162.8
Observations	10,115,792	14,691,136	9,244,829	12,818,625
Panel B: by destination's pollution level				
AQI difference	0.059***	0.032***	0.058***	0.040**
	(0.011)	(0.007)	(0.019)	(0.017)
Coefficient difference		0.026***		0.018**
p-value		0.000		0.033
Benchmark	132.861	144.121	149.287	162.718
Relative effect	2.209	1.117	1.946	1.224
KP F stat.	175.3	172.6	173.7	158
Observations	10,002,242	14,804,702	9,354,356	12,709,115
CityPair FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes

Notes. This table presents heterogeneity in pollution level of origin or destination cities. Panel A report the heterogeneity in origin's pollution level, and Panel B reports the heterogeneity in origin's pollution level. Columns (1)–(2) and (3)–(4) report the results for ticket bookings leaving in 0–1 days and 2–7 days, respectively. For each column, the dependent variable is the probability of individuals from an origin city i choosing to travel to destination city j on reservation date t in a specific booking period. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level.

Table 8
Refund and cancellation.

	(1)	(2)	(3)
	Ratio of refunds/cancellations		
	Train	Flight	
	Refund	Refund	Cancellation
Panel B: 2SLS			
AQI difference	0.006**	0.003	0.001
	(0.002)	(0.009)	(0.003)
Benchmark	18.434	42.057	25.556
Relative effect (%)	1.641	0.382	0.148
KP F stat.	274.2	154.7	154.7
Panel A: OLS			
AQI difference	−0.001	0.001	−0.001
	(0.001)	(0.004)	(0.001)
Benchmark	18.434	42.057	25.556
Relative effect (%)	−0.165	0.066	−0.227
Observations	30,574,889	8,356,476	8,356,476
CityPair FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes

Notes. This table presents the effect of the AQI gap on ticket refunds/cancellations. Panel A reports OLS estimates, and Panel B reports 2SLS estimates using equations (5) and (6). The dependent variable of columns (1) is the refund rate of train tickets and that of columns (2)–(3) is the refund/cancellation rate of airline tickets. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

higher than for train tickets (almost zero) and individuals are more cautious when purchasing airline tickets.

What role does official forecasting play in the decision-making process of trip-booking? If people were rational, today's pollution level should have little predictive power on tomorrow's travel, conditional on the forecasted pollution level tomorrow.[24] To explore this hypothesis, we collect pollution forecast data for 46 cities from September 2023 to January 2024 from the Air Quality Forecasting Information Release System.[25] We conduct an extended analysis, incorporating pollution forecasts for both the origin and destination cities into our regression models. Specifically, for travels departing within a 0–1 day window, we utilize the 24-h AQI forecast, and for travels departing in 2–7 days, we refer to the 72-h AQI forecast. The upwind AQI difference is used as an instrumental variable for AQI difference at the booking date, as in our baseline approach.

Table A.10 presents suggestive evidence that, conditional on AQI forecasts, the AQI difference on the booking date has a negligible impact on travel flows between a city pair. This suggests that travelers make trip decisions based on these forecasts to avoid future pollution, rather than merely reacting irrationally to current pollution. However, it is important to note that this discussion is based on a relatively short sample period of five months and that the air quality in China has been improving over time.[26] Interestingly, it is AQI forecasts in the origin city instead of that in the destination city that have a noticeable impact on travel flows. This aligns with the results showcased in Table 6, which indicates that higher pollution levels in the origin city has a larger effect.

Review the findings we have so far: 1) Train tickets booked on days with a higher AQI-gap are more likely to be refunded later, indicating behavioral factors in trip-booking behavior; 2) Over a small sample of five months, conditional on pollution forecasts, today's AQI difference has little predictive power on future travel, indicating rational factors in trip-booking (but this finding should be interpreted with caution due to the limited sample). Overall, these findings suggest that both rational and behavioral factors play a role in the decision-making process for booking trips to avoid air pollution.

5.2. Intertemporal substitution

To test whether these "pollution-induced" trips are newly generated or just intertemporal shifts—individuals bringing forward the travel plans that they have already made when the origin city is more polluted—we estimate the cumulative effects over different time periods. Under the distributed lag model $NumRatio_{ijtr} = \sum_{\tau=0}^{k} \beta_\tau (P_{i,t-\tau} - P_{j,t-\tau}) + f(W_{it} - W_{jt}) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijtr}$, we are concerned about high autocorrelation among the lagged terms $P_{i,t-\tau} - P_{j,t-\tau}$, so we follow Barwick et al. (2024) and use the IV version of a flexible distributed lag

[24] We thank one of the referees for proposing the test to investigate the role of official pollution forecasts. By "rational", we mean that individuals make trip-booking decisions with the intention of gaining air quality improvements through trips, based on the information available to them.

[25] The website we collected data from is https://air.cnemc.cn:18014/.

[26] Specifically, the average daily AQI during our sample period was 58.6, compared to an average of 66.6 during the years 2017–2019.

model, which allows for flexible and smooth long-term effects and deals with the high-autocorrelation issues. β_τ are specified as cubic B-spline functions of time with one segment:

$$\beta_\tau = \gamma_0 + \gamma_1\tau + \gamma_2\tau^2 + \gamma_3\tau^3. \quad (7)$$

The contemporaneous effect is $\beta_0 = \gamma_0$, and the effect of pollution from τ days in the past is $\beta_\tau = \gamma_0 + \gamma_1\tau + \gamma_2\tau^2 + \gamma_3\tau^3$. Then, we substitute β_τ into the distributed lag model and get

$$\begin{aligned} NumRatio_{ijtr} = {} & \sum_{\tau=0}^{k}\beta_\tau\left(P_{i,t-\tau} - P_{j,t-\tau}\right) + f\left(W_{it} - W_{jt}\right) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijt} = \gamma_0\left(P_{it} - P_{jt}\right) + \left(\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3\right)\left(P_{i,t-1} - P_{j,t-1}\right) + \ldots \\ & + \left(\gamma_0 + \gamma_1 k + \gamma_2 k^2 + \gamma_3 k^3\right)\left(P_{i,t-k} - P_{j,t-k}\right) + f\left(W_{it} - W_{jt}\right) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijtr} = \gamma_0 \nu_{1,ijt} + \gamma_1 \nu_{2,ijt} + \gamma_2 \nu_{3,ijt} + \gamma_3 \nu_{4,ijt} \\ & + f\left(W_{it} - W_{jt}\right) + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijtr}, \end{aligned} \quad (8)$$

where $\nu_{1,ijt} = \left(P_{it} - P_{jt}\right) + \left(P_{i,t-1} - P_{j,t-1}\right) + \ldots + \left(P_{i,t-k} - P_{j,t-k}\right)$, $\nu_{2,ijt} = \left(P_{i,t-1} - P_{j,t-1}\right) + 2\left(P_{i,t-2} - P_{j,t-2}\right) + \ldots + k\left(P_{i,t-k} - P_{j,t-k}\right)$, $\nu_{3,ijt} = \left(P_{i,t-1} - P_{j,t-1}\right) + 2^2\left(P_{i,t-2} - P_{j,t-2}\right) + \ldots + k^2\left(P_{i,t-k} - P_{j,t-k}\right)$, $\nu_{4,ijt} = \left(P_{i,t-1} - P_{j,t-1}\right) + 2^3\left(P_{i,t-2} - P_{j,t-2}\right) + \ldots + k^3\left(P_{i,t-k} - P_{j,t-k}\right)$, respectively. We can then estimate the cumulative effect $\sum_{\tau=0}^{k}\beta_\tau$, which is a linear combination of $\{\gamma_i\}_{i=0}^{3}$. If the positive coefficients in the baseline results are entirely caused by intertemporal substitution over a period, the cumulative effect $\sum_{\tau=0}^{k}\beta_\tau$ should be insignificantly different from zero.

Fig. 6 plots the path of cumulative effects over different time periods, with each dot denoting the sum of contemporaneous and lagged effects. We can see the cumulative effect increases over the seven-day window. In particular, a 50-unit AQI gap leads to a 1.99% (3.13%) increase in overall ticket bookings leaving within zero to one (2–7) days, compared with the contemporaneous effect of 0.94% (0.93%). We find no evidence of intertemporal shifting within a week for train tickets, which means the pollution-induced train trips are newly generated; however, we find the cumulative effects within a week for airline tickets mute, which means the airline tickets bookings are likely to be intertemporal shifting.

To study how the effect of pollution on ticket bookings changes over time, we also plot the coefficients of the lagged pollution gaps in Appendix Figure A.4. We find for travels by train, today's booking decisions are also affected by the pollution gap in the past few days, but to a lesser extent than by pollution today; in contrast, airline ticket bookings are not affected by the pollution gap in the past few days.

We also estimate the lead effects in a similar manner. It is important to note that we use pollution gap on the *booking* date instead of the departure date, so the existence of lead effects indicates that individuals respond to future pollution levels to avoid exposure. As shown in Appendix Figure A.5 Panel (a), for trips departing within 0–1 days, we find that the pollution gap on day 0 has the greatest effect. Additionally, there are significant lead effects at t = 1 (one day after the booking date), with coefficients approximately half the magnitude of those at t = 0. Lead effects from t = 2 to t = 7 are much smaller although statistically significant. In contrast, for trips departing within 2–7 days, as shown in Panel (b) of Figure A.5, booking decisions respond to future pollution gaps over four days, with the coefficients decreasing in magnitude from day 0 to day 4.

Last but not least, the effects of pollution gap on the booking date (t = 0) are the largest in specifications that control a series of lagged or lead terms, as shown in Figure A.4 and Figure A.5. This further supports our theory that individuals' trip decisions are directly affected by pollution levels on the booking date.

5.3. Willingness to pay

Lastly, we use the instrumental variable approach to estimate the willingness to pay for short-term trips to avoid air pollution. Exploiting the price data of airline and train tickets, we can firstly estimate the cost of pollution in the following reduced-form analysis:

$$NumRatio_{ijtr} = \alpha_0 + \alpha_1\left(P_{it} - P_{jt}\right) + \alpha_2 price_{ijtr} + \gamma_{im} + \theta_{jm} + \delta_{ij} + \phi_t + \varepsilon_{ijtr}, \quad (9)$$

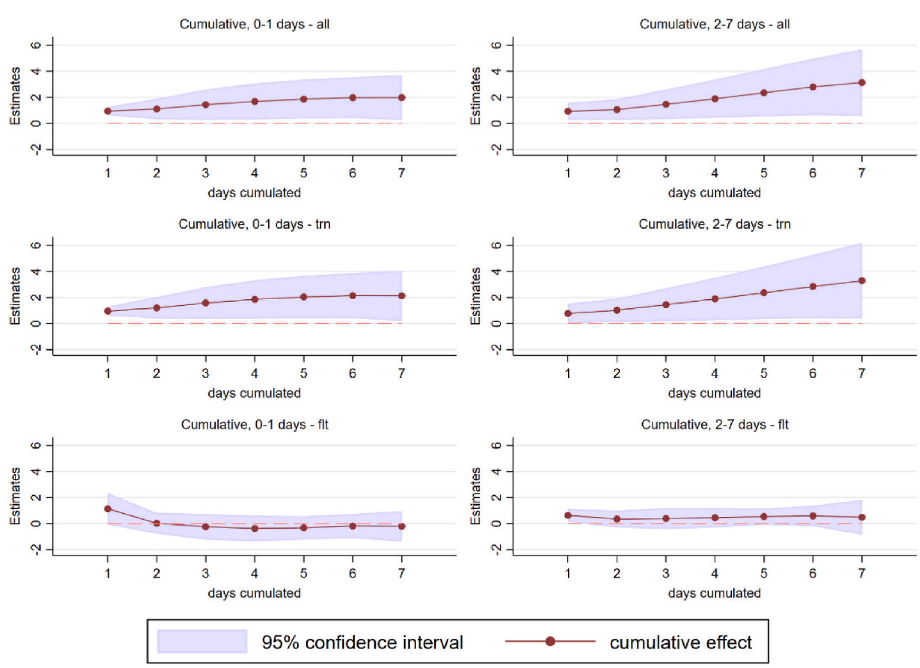


Fig. 6. Cumulative Effects over Different Periods
Notes. This figure plots the relative cumulative effects and the 95% confidence intervals over different time periods. The figures on the left (right) show the results for tickets booked to depart in 0–1 days (2–7 days). Each dot denotes the sum of relative contemporaneous effects and the lagged effects estimated using equation (8), which is $\sum_{\tau=0}^{k}\beta_\tau/benchmark$; for example, the first dot in each figure denotes the relative effect of today's and yesterday's pollution.

Table 9
Willingness to pay.

	(1)	(2)	(3)	(4)
	0–1 days		2–7 days	
Panel A: 2SLS				
1st stage	AQI difference	Price	AQI difference	Price
Upwind AQI difference	1.687***	0.065***	1.663***	0.030**
	(0.110)	(0.020)	(0.112)	(0.013)
# of trains available	−0.001	−1.649***	0.009	−2.055***
	(0.008)	(0.226)	(0.009)	(0.256)
# of flights available	−0.003	25.393***	−0.006	20.313***
	(0.030)	(4.088)	(0.028)	(2.932)
2nd stage	Dependent variable: Ratio of ticket bookings			
AQI difference	0.041***		0.064***	
	(0.007)		(0.016)	
Price	−0.133		−0.644***	
	(0.095)		(0.196)	
Willingness to pay (Yuan)	0.31		0.10	
First-stage Sanderson-Windmeijer F-test:				
AQI difference	139.61		129.99	
p-value	0.000		0.000	
Price	40.05		46.38	
p-value	0.000		0.000	
Kleibergen-Paap (KP) F statistic	25.96		30.29	
Panel B: OLS	Dependent variable: Ratio of ticket bookings			
AQI difference	0.002		0.011***	
	(0.002)		(0.004)	
Price	−0.016***		−0.037***	
	(0.004)		(0.006)	
Willingness to pay (Yuan)	0.14		0.31	
Observations	24,806,951		22,063,482	
CityPair FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes

Notes. This table presents estimates of willingness to pay for clean air. Panels A and B report the 2SLS and OLS estimates, respectively. Price is the overall average price of traveling between a city pair. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *Willingness to pay* is the absolute value of the ratio between the estimated coefficient of AQI difference and that of price. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01.

where $price_{ijtr}$ is the average price (incorporating flight and train) of trips from city i to city j on date t in the reservation window r. We use the marginal rate of substitution between the air-quality difference and ticket price as the measure of WTP, which is $-\frac{\alpha_1}{\alpha_2}$ from the reduced-form analysis. The underlying assumption is that price is not correlated with the error term (ε_{ijt}) after controlling for $\gamma_{im}, \theta_{jm}, \delta_{ij}$ and ϕ_t.

Nevertheless, the *price* would be positively correlated with the number of trips in the city-pair-date cell, because of the demand-side shock, which will lead to a rather smaller absolute value of α_2. Our estimates for the coefficient of price would thus have an upper bias from the actual value, and this problem is more severe for trips leaving in two to seven days. To mitigate this concern, we instrument the price from the supply-side variation, that is, the number of flights and trains available between the city pair on the reservation date, for each city-pair-reservation-date cell.[27] The overall average price of a trip incorporates both train and airline ticket prices, and the average price reflects the substitution between trips by train and airline. The price of train tickets between two cities is relatively stable over time and has variations between high-speed trains and normal trains, and first-class and second-class. The price of airline tickets is determined by the market in time; however, it is also affected by the availability of trains because of the competition effect (Fang et al., 2020; Zhang et al., 2019). Thus, the overall average price is correlated with the change in the number of available trains and the number of available flights even after controlling for city-pair fixed effects, which meets the relevance condition. In addition, the number of trains and flights on a particular date for a given city pair is planned by the Ministry of Railway and the airlines. Although airline companies consider the demand side when arranging flights, the arrangements are usually based on historical data in a broader date range. This channel would be absorbed by city-pair fixed effects and city-month fixed effects. Moreover, any demand shock within the multi-level fixed effects that the flight company could not anticipate when arranging flights doesn't invalidate the IV's exogenous condition. Overall, we believe the variation on the supply side across dates within the city pair would be a valid IV for the price.

Table 9 reports the 2SLS estimates in Panel A and OLS estimates in Panel B, respectively. The OLS estimation shows the average WTP of an individual departing within one day (2–7 days) is 0.14 (0.31) yuan for a one-unit improvement in AQI.[28] That the WTP for trips departing in a longer period is larger contradicts the common sense, as people would pay more for urgent trips. The reason is that the price in longer periods has a larger variation and stronger correlation with unobservable demand shock. We further use instrument variables to predict the AQI gap and price together in the first stage. As expected, the coefficient of *price* increases after exploiting IVs for price. The Sanderson and Windmeijer (2016) first-stage F statistics for each of the endogenous variables are presented in Table 9 and are much larger than conventionally acceptable thresholds, which reject the null of weak instruments on the two endogenous variables. The 2SLS estimation shows that an individual who departs within one day (2–7 days), on average, is willing to pay

[27] In particular, we use the number of distinct trains/flights on the departure date. For example, if flight CX983 from city i to city j makes three trips in 2–7 days, we only treat it as one choice in 2–7 days.

[28] The WTP estimated using $PM_{2.5}$ is similar to AQI.

0.31 (0.10) yuan for a one-unit improvement in AQI. In other words, an individual would be willing to pay 0.31 yuan for each unit improvement in AQI (avoiding one unit AQI) within one day, and 0.10 yuan for each unit improvement in AQI for avoiding pollution within 2–7 days. Regarding the IV results, the WTP for urgent trips is larger than that of trips departing in a longer period. What is the magnitude of WTP for the short-term pollution avoidance? In our data, an individual travels 3.61 times per year, on average, which means an average individual leaving in zero to one days (2–7 days) is willing to pay 1.12 (0.36) yuan annually for a one-unit improvement in AQI. Therefore, the WTP of a household of two to three is about 2.24–3.36 (0.72–1.08) yuan for one unit of AQI in a trip leaving in zero to one days (2–7 days), respectively. Our estimates are smaller than the WTP estimated from air purifier purchases by Ito and Zhang (2020) and much smaller than the WTP estimated from long-term migration by Bayer et al. (2009) and Freeman et al. (2019),[29] which is reasonable because air purifiers are durable goods and migration is a long-term decision, whereas travel is transient consumption. In addition, we use their final decisions (net of cancellation) instead of the initial decisions to measure WTP. If using the initial decisions, the WTP would be larger. Our estimate is likely to be a lower bound because we only consider the ticket costs of the pollution-induced trip and do not incorporate other possible costs such as accommodation costs. Our estimation of WTP is not a substitute estimation for Ito and Zhang (2020) but rather a complementation. Besides the preference for better *indoor* air quality, we want to emphasize that people may also need short trips to enjoy better *outdoor* air quality, which is mainly for mental health.

6. Conclusion

In this paper, we study how air-pollution differences can affect the short-term travel flow between cities. Utilizing unique data from one of the largest flight and train ticket bookings providers in China, we can estimate how the travel flows change with the air pollution on the *dates that the travel decisions are made*, and investigate how the effects vary with demographic characteristics such as age. We find that a 50-unit difference in AQI between a city pair leads the travel flow to the cleaner city to increase by 1.30%–1.33%. Additionally, middle-aged people are more responsive to the air-pollution difference than children and the old. Our analysis suggests people are more likely to travel by train to intra-province cities with more tourist attractions.

This paper also sheds light on the behavioral factors that play a role in the decision-making process, by showing an increase in the refund rate of those pollution-induced ticket bookings. Our estimates on the marginal rate of substitution between air-quality difference and ticket price also add to a growing body of work on measuring willingness to pay for clean air.

These pollution-induced trips can bring indirect costs, such as additional greenhouse gas emissions and loss of tourism, which may offset part of the health benefits. We try to calculate the additional greenhouse gas emissions due to the pollution-induced trips following Lin et al. (2021). The detailed calculation steps are described in Appendix A. The back-of-envelope calculation shows the increased travel due to air pollution leads to an 8,292.45-ton increase in CO2e annually—we need to plant 460,692–2,073,114 more trees to absorb these additional GHG emissions.[30]

CRediT authorship contribution statement

Ruochen Dai: Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Writing – review & editing. **Dongmei Guo:** Conceptualization, Data curation, Methodology, Supervision, Writing – review & editing. **Yajie Han:** Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Yu Qin:** Conceptualization, Data curation, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Data availability

The authors do not have permission to share data.

Appendix

A. Greenhouse Gas Emission

Following Lin et al. (2021), we first calculate the greenhouse gas emission (GHG) by different transportation modes, namely, train and airline, for each city pair on each day and then aggregate. Using the estimates from the nonlinear specification in Table A.2, the annual GHG is calculated as follows:

$$GHS\_annual = \frac{1}{3}\sum_{ij,t}\sum_{h}\sum_{x=2}^{6}\delta_h \times Dis_{ij,h} \times Num_{ijt} \times \frac{\beta_i \times D[AQI\_difference\_xtile_x = 1]}{Benchmark_h} \tag{10}$$

where h denotes transportation mode, that is, high-speed railway, traditional train, and airplane. The emission factor (δ_h) for high-speed railway, traditional train, and airplane are 0.04, 0.006, 0.018 kg CO2e per passenger*km, respectively. For train, the distances between a city pair by transportation mode ($Dis_{ij,h}$) are imputed using the train ticket price[31]; for airplane, the distance is calculated as the distance between two cities' airports. $D[AQI\_difference\_xtile_x = 1]$ denotes the dummy when the AQI difference is in xtile x; Num_{ijt} is the number of people traveling from city i to city j on date t. $Benchmark_h$ is the mean value of the dependent variable in equation (1) in the group (AQI_difference <35) for each transportation mode h. Because our sample is from 2017 to 2019, we derive the annual GHG emission by multiplying $\frac{1}{3}$.

[29] Ito and Zhang (2020) estimate that a household's *annual* WTP for 1 $\mu g/m^3$ improvement of PM_{10} is about \$1.34 (about 9 yuan). Bayer et al. (2009) estimate that the median household in U.S. would pay \$149–\$185 for a one-unit reduction in PM_{10} concentrations, in constant 1982–1984 dollars; Freeman et al. (2019) estimate that a median household in China is willing to pay \$21.70 for a one-unit decline in annual $PM_{2.5}$ concentration.

[30] According to the statistics released by National Forestry and Grassland Administration, a tree can absorb and store 4–18 kg of CO_2 per year.

[31] The average price for traditional train is 0.31 yuan/km, and the average price for high-speed railway is 0.46 yuan/km.

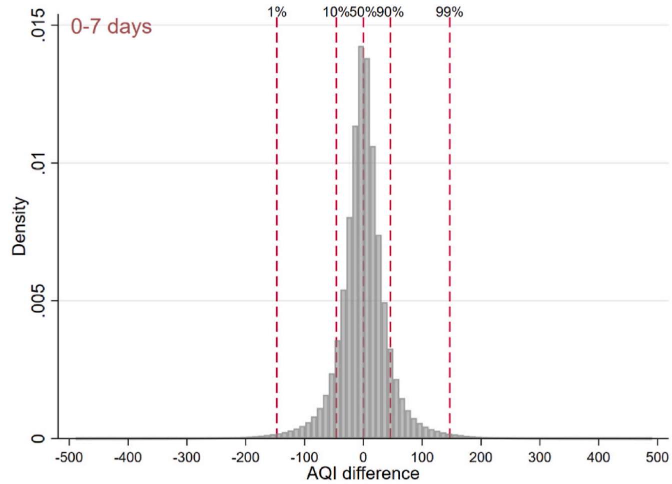
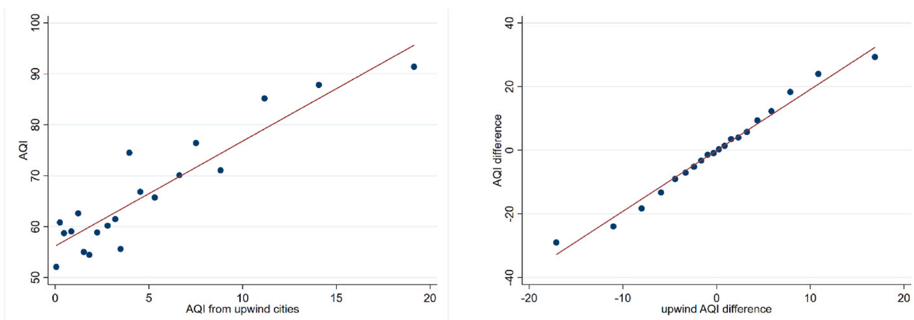


Fig. A.1. Distribution of AQI Difference
Notes. This figure shows the distribution of the AQI difference. The dotted lines denote the 1st, 10th, 50th, 90th, and 99th percentiles of AQI difference, respectively.



Panel A: city level Panel B: city-pair level

Fig. A.2. Correlation between the Endogenous Variable and the IV

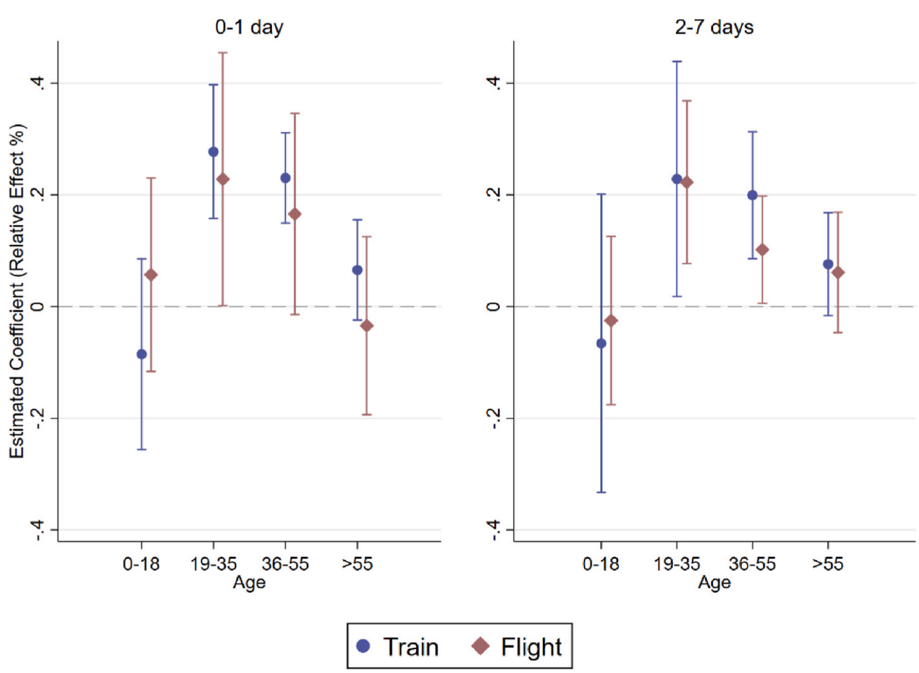


Fig. A.3. Heterogeneity in Age, by Train and Airplane
Notes. This figure plots the relative effects and the 95% confidence intervals for different age cohorts booking train or airline tickets. The figure on the left (right) shows the results for tickets booked to depart within a day (2–7 days). Each dot denotes the relative effect (%) of a 50-unit increase in the AQI gap, which is calculated by dividing the estimated coefficient (reported in Table A.7) of the AQI difference by the benchmark value of the dependent variable and then multiplying by 50.

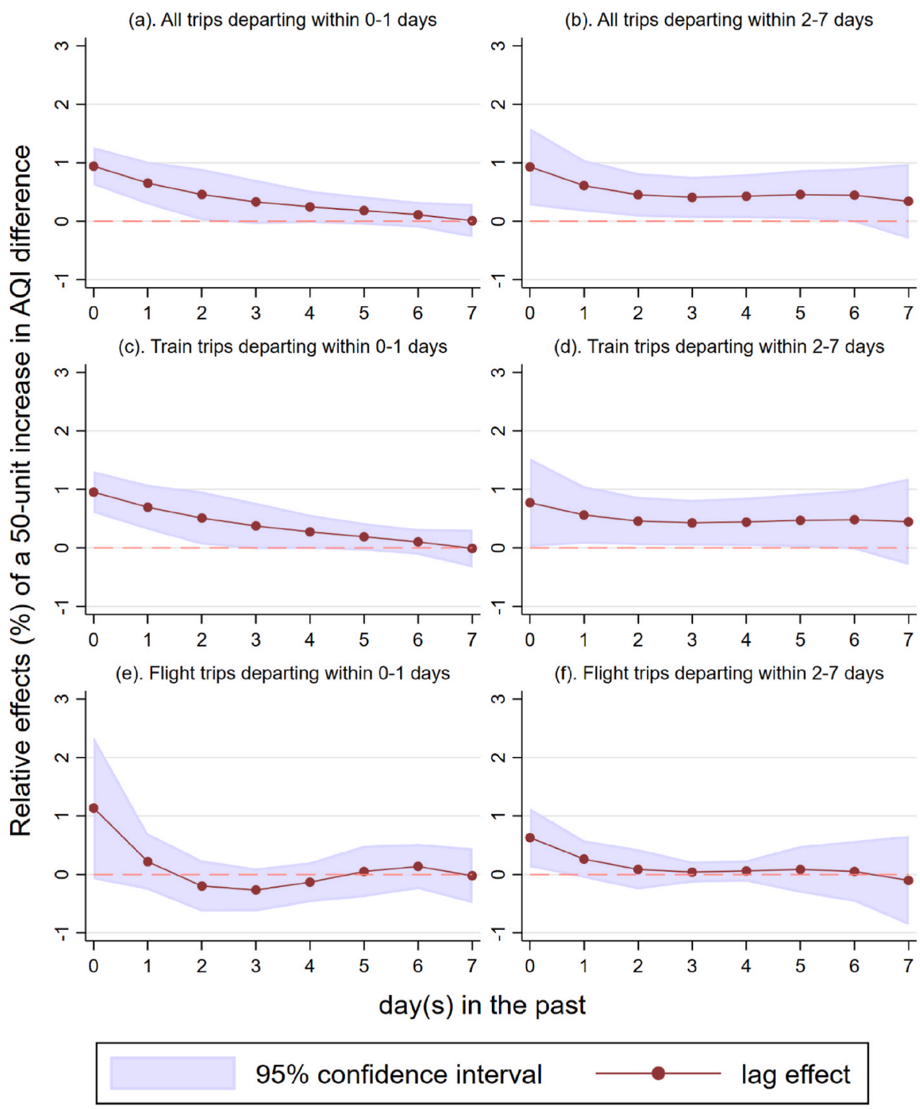


Fig. A.4. Lagged Effects of AQI Difference

Notes. This figure plots the relative lagged effects of the air-pollution gap, denoted by $\{\beta_\tau\}_{i=1}^{7}/benchmark*50$, as estimated using the flexible distributed lag model in equation (8). Lag effects indicate that booking decisions respond to pollution gaps in the past. Panel (a) and (b) present the effects for all trips, while Panels (c)–(d) and Panels (e)–(f) respectively present the effects for train and flight trips. The three subfigures in the left depict the effects on trips departing within 0–1 days, while the three subfigures in the right illustrate the effects on trips departing within 2–7 days. Each dot denotes the estimated relative effect, with the accompanying blue area indicating the 95% confidence interval. The x-axis corresponds to the number of days preceding the present day; for example, x = 1 refers to yesterday, and x = 2 denotes two days before the current date.

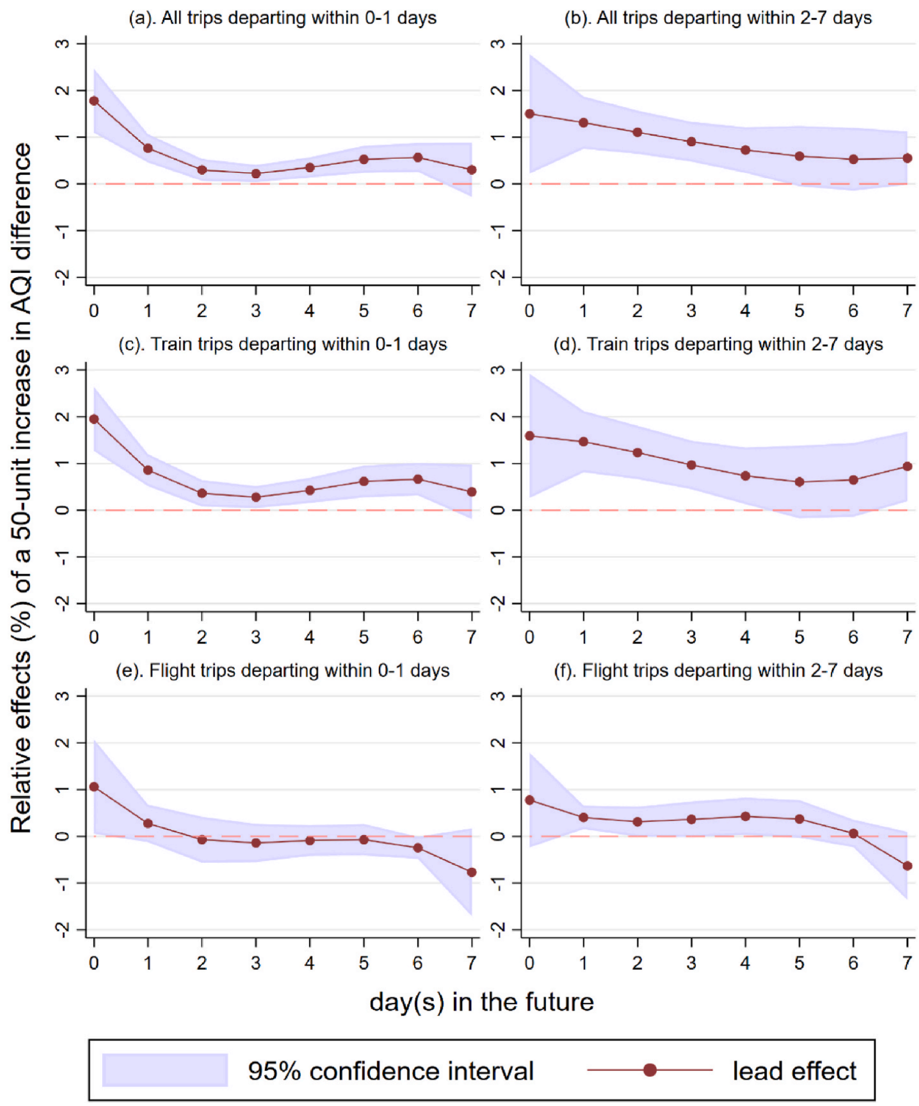


Fig. A.5. Lead Effects of AQI Difference

Notes. This figure plots the relative lead effects of the air-pollution gap, denoted by $\{\beta_\tau\}_{i=-7}^{-1}/benchmark*50$, as estimated using the flexible distributed lag model in equation (8). Leads effects indicate that current booking decisions respond to future pollution gaps. Panel (a) and (b) present the effects for all trips, while Panels (c)–(d) and Panels (e)–(f) respectively present the effects for train and flight trips. The three subfigures in the left depict the effects on trips departing within 0–1 days, while the three subfigures in the right illustrate the effects on trips departing within 2–7 days. Each dot denotes the estimated relative effect, with the accompanying blue area indicating the 95% confidence interval. The x-axis corresponds to the number of days after the current day; for example, x = 1 refers to one day after the booking date, and x = 2 denotes two days after the current date.

Table A.1
Platform Choice during the Shopping Festival

	(1)	(2)	(3)	(4)	(5)	(6)
	Ratio of ticket bookings					
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
AQI difference	0.036***	0.042***	0.038***	0.040**	0.081*	0.050***
	(0.006)	(0.013)	(0.008)	(0.017)	(0.042)	(0.018)
AQI difference*the week of 11.11	−0.006	−0.004	−0.019	0.025	0.190	0.022
	(0.014)	(0.022)	(0.017)	(0.034)	(0.180)	(0.049)
Observations	24,806,951	22,063,482	22,292,946	18,608,699	5,355,666	6,541,710
Benchmark	139.582	157.03	149.35	179.00	372.12	304.99
Relative effect of AQI difference (%)	1.30	1.34	1.27	1.13	1.09	0.82
KP F stat.	114.9	109.1	117.3	111.5	73.43	76.49
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents indirect test of the platform choice independence assumption by utilizing the shock of Double Eleven. *AQI difference*the week of 11.11* is instrumented by the interaction of upwind pollution difference and the Double-Eleven week. We include all the pairwise terms. For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.2
Non-linearity

	(1)	(2)	(3)	(4)	(5)	(6)
	Ratio of ticket bookings					
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
[35, 50)	0.320	0.931**	0.280	1.161**	2.531**	1.763
	(0.267)	(0.393)	(0.311)	(0.550)	(1.249)	(1.106)
[50, 75)	1.912***	3.016***	2.070***	3.872***	0.807	1.963**
	(0.467)	(0.585)	(0.535)	(0.810)	(1.339)	(0.910)
[75, 150)	5.431***	5.159***	6.076***	6.553***	3.918*	4.197***
	(0.816)	(0.838)	(0.902)	(1.110)	(2.246)	(1.179)
[150, 250)	9.541***	6.744***	11.404***	8.752***	6.419**	4.618**
	(1.255)	(1.158)	(1.465)	(1.623)	(3.111)	(2.025)
[250)	14.416***	7.448***	22.259***	10.676***	−0.968	−0.068
	(2.118)	(1.654)	(5.316)	(1.591)	(3.617)	(2.654)
Observations	24,806,951	22,063,482	22,292,946	18,608,699	5,355,666	6,541,710
R-squared	0.831	0.694	0.835	0.666	0.744	0.796
Benchmark	150.546	166.052	160.770	189.377	386.784	315.350
Relative effect (%)	9.576	4.485	13.845	5.638	−0.250	−0.022
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the nonlinear impact of air pollution on travel decisions. We decompose continuous AQI into six bins. For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient of the last bin and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.3
Robustness – Std. Err. One-way Clustered at City-pair

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
	Dependent variable: AQI difference					
Panel A: 2SLS						
1st stage						
Upwind AQI difference	1.687***	1.663***	1.671***	1.634***	1.684***	1.695***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.015)	(0.014)
KP F stat.	46951	42952	43476	38361	12669	14280
	Dependent variable: Ratio of ticket bookings					
2nd stage						
AQI difference	0.036***	0.042***	0.038***	0.041***	0.083***	0.050***
	(0.004)	(0.005)	(0.004)	(0.006)	(0.018)	(0.011)
Benchmark	139.582	157.025	149.355	178.998	372.117	304.989
Relative effect (%)	1.301	1.334	1.263	1.134	1.117	0.828
	Dependent variable: Ratio of ticket bookings					
Panel B: OLS						
AQI difference	0.002	0.011***	0.004**	0.016***	0.017**	0.007*
	(0.001)	(0.002)	(0.002)	(0.003)	(0.007)	(0.004)
Observations	24,806,951	22,063,482	22,292,946	18,608,699	5,355,666	6,541,710
Benchmark	139.582	157.025	149.355	178.998	372.117	304.989
Relative effect (%)	0.076	0.359	0.124	0.451	0.232	0.117
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the robustness check of clustering the standard errors at the city-pair level. Panel A reports the 2SLS estimates of equations (3) and (4), and Panel B reports the OLS estimates of equation (1). For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.4
Robustness – IV Radius 150 km–350km

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
	Dependent variable: AQI difference					
Panel A: 2SLS						
1st stage						
Upwind AQI difference	1.953***	1.947***	1.904***	1.865***	2.177***	2.200***
	(0.094)	(0.100)	(0.089)	(0.094)	(0.166)	(0.161)
KP F stat.	435.3	379.9	458.6	393	172.9	186.9
	Dependent variable: Ratio of ticket bookings					
2nd stage						
AQI difference	0.057***	0.070***	0.062***	0.079***	0.105***	0.089***
	(0.010)	(0.020)	(0.012)	(0.025)	(0.039)	(0.031)
Benchmark	139.684	157.161	149.544	179.345	373.784	305.811
Relative effect (%)	2.045	2.212	2.067	2.208	1.409	1.453
	Dependent variable: Ratio of ticket bookings					
Panel B: OLS						
AQI difference	0.002	0.011***	0.004	0.016**	0.016	0.007**
	(0.002)	(0.004)	(0.003)	(0.007)	(0.021)	(0.003)
Benchmark	139.684	157.161	149.544	179.345	373.784	305.811
Relative effect (%)	0.073	0.363	0.122	0.456	0.215	0.115
Observations	24,970,659	22,224,759	22,438,128	18,736,009	5,391,423	6,598,087
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the robustness check of changing the radius (within 150–350 km) used to construct the instrumental variable in Fig. 1. Panel A reports the 2SLS estimates of equations (3) and (4), and Panel B reports the OLS estimates of equation (1). For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.5
Robustness – Using Different Denominators in the Dependent Variable.

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
Panel A: dependent variable: ratio2 = num/daily average trips of city *i* over the sample period						
AQI difference	0.037***	0.044***	0.037***	0.043**	0.088*	0.049**
	(0.006)	(0.014)	(0.008)	(0.017)	(0.045)	(0.019)
Relative effect (%)	1.313	1.380	1.228	1.196	1.158	0.779
Observations	24,806,951	22,063,482	22,292,946	18,608,699	5,355,666	6,541,710
Panel B: dependent variable: ratio3 = num/daily average of trips of city *i* in 2017						
AQI difference	0.055***	0.052**	0.053***	0.051**	0.105**	0.066***
	(0.012)	(0.021)	(0.014)	(0.026)	(0.049)	(0.025)
Relative effect (%)	1.427	1.197	1.279	1.011	1.074	0.812
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the robustness check using alternative dependent variables. Panel A reports the 2SLS estimates of using ratio2, with the denominator being the average daily number of trips over the three-year period (2017–2019). Panel B presents the results using ratio3, with the denominator being the average daily number of trips in the first year of the sample (2017). Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions for control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.6
Robustness – Logarithm of the Number of Trips as the Dependent Variable.

	(1)	(2)	(3)	(4)	(5)	(6)
	All trips leaving in		Train trips leaving in		Flight trips leaving in	
	0–1 days	2–7 days	0–1 days	2–7 days	0–1 days	2–7 days
	Dependent variable: AQI difference					
Panel A: 2SLS						
1st stage	1.68686***	1.66296***	1.67110***	1.63394***	1.68403***	1.69526***
Upwind AQI difference	(0.11038)	(0.11197)	(0.10788)	(0.10816)	(0.13892)	(0.13716)
KP F stat.	233.6	220.6	240	228.2	146.9	152.8
	Dependent variable: ln(number of trips)					
2nd stage						
AQI difference	0.00009***	0.00008**	0.00009***	0.00010**	0.00006*	0.00004
	(0.00002)	(0.00004)	(0.00003)	(0.00005)	(0.00003)	(0.00003)
Effect of a 50-unit increase in AQI difference (%)	0.471	0.414	0.475	0.491	0.296	0.193
	Dependent variable: ln(number of trips)					
Panel B: OLS						
AQI difference	0.00001*	0.00003**	0.00001	0.00004*	0.00002	0.00001
	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00001)	(0.00001)
Effect of a 50-unit increase in AQI difference (%)	0.069	0.159	0.056	0.192	0.086	0.038
Observations	24,806,951	22,063,482	22,292,946	18,608,699	5,355,666	6,541,710
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the robustness check of using the logarithm of the number of trips as an alternative dependent variable. Panel A reports the 2SLS estimates of equations (3) and (4), and Panel B reports the OLS estimates of equation (1). Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings, respectively. Columns (1), (3), and (5) show the results for tickets booked to depart within a day, and columns (2), (4), and (6) report the results for tickets booked to depart in 2–7 days. All regressions for control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Effect of a 50-unit increase in AQI difference (%)* is calculated using $\left(e^{\beta\times 50}-1\right)\times 100$. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.7
Heterogeneity in Age.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ratio of ticket bookings							
	0–1 days				2–7 days			
	0–18 y/o	19–35 y/o	36–55 y/o	>55 y/o	0–18 y/o	19–35 y/o	36–55 y/o	>55 y/o
Panel A: All								
AQI difference	−0.022	0.046***	0.047***	0.024*	−0.028	0.053***	0.050***	0.040**
	(0.053)	(0.009)	(0.007)	(0.015)	(0.067)	(0.019)	(0.012)	(0.017)
Observations	5,321,255	19,759,268	17,932,479	7,301,197	5,730,427	17,354,573	15,120,266	7,510,598
Benchmark	646.891	175.189	193.132	472.761	600.134	199.593	229.124	459.893
Relative effect (%)	−0.170	1.321	1.205	0.255	−0.235	1.330	1.101	0.436
KP F stat.	136	212.6	207.4	160.9	142.5	200.7	198.5	169.3
Panel B: Train								
AQI difference	−0.064	0.052***	0.049***	0.036	−0.055	0.053**	0.056***	0.046
	(0.066)	(0.012)	(0.009)	(0.025)	(0.113)	(0.025)	(0.016)	(0.028)
Observations	4,397,963	17,647,868	15,595,589	5,996,948	3,966,195	14,399,804	11,927,236	5,470,789
Benchmark	751.930	188.628	213.502	552.960	829.321	231.277	279.068	603.433
Relative effect (%)	−0.427	1.387	1.151	0.328	−0.330	1.142	0.997	0.380
KP F stat.	133.5	218.2	212.2	159.7	134.9	209.2	205.1	165.3
Panel C: Flight								
AQI difference	0.087	0.107**	0.076*	−0.036	−0.021	0.084***	0.038**	0.042
	(0.136)	(0.054)	(0.042)	(0.084)	(0.065)	(0.028)	(0.018)	(0.038)
Observations	1,270,091	4,232,763	4,309,267	1,893,607	2,325,456	5,259,480	5,275,835	2,898,440
Benchmark	1533.688	470.773	461.159	1037.135	844.299	378.248	377.572	682.838
Relative effect (%)	0.285	1.141	0.828	−0.172	−0.126	1.113	0.509	0.305
KP F stat.	119.2	143.9	147.7	141.4	129.9	147.9	153.8	148.5
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the heterogeneity in age by dividing the sample into four groups: 0–18, 19–35, 36–55, and more than 55 years old. Panels A, B, and C report the 2SLS estimates of equations (3) and (4) for all, train, airline tickets respectively. For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for all, train, and airline ticket bookings respectively. Columns (1)–(4) show the results for tickets booked to depart within a day, and columns (5)–(8) report the results for tickets booked to depart in 2–7 days. All regressions control for linear and quadratic forms of weather differences between origin and destination city, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.8
Heterogeneity – Intra-province and Inter-province Trips.

	(1)	(2)	(3)	(4)
	Ratio of ticket bookings			
	Train		Flight	
	Inter-prov.	Intra-prov.	Inter-prov.	Intra-prov.
Panel A: 0–1 days				
AQI difference	0.013***	0.737***	0.076*	0.351
	(0.004)	(0.204)	(0.046)	(0.451)
Observations	19,103,145	3,189,794	5,072,038	283,078
Benchmark	64.606	656.901	331.814	1090.622
Relative effect (%)	0.995	5.608	1.143	1.608
KP F stat.	242.1	130.8	144.2	72.99
Coefficient difference		−0.724***		−0.275***
p-value		0.000		0.000
Panel B: 2–7 days				
AQI difference	0.025***	0.542	0.041**	0.422
	(0.007)	(0.409)	(0.018)	(0.418)
Observations	15,815,557	2,793,129	6,217,671	323,707
Benchmark	104.594	600.281	280.467	771.005
Relative effect (%)	1.176	4.513	0.729	2.737
KP F stat.	232.4	122	150.4	70.76
Coefficient difference		−0.517***		−0.381***
p-value		0.000		0.000
CityPair FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes

Notes. This table presents heterogeneity in destination cities. Panels A and B report the 2SLS estimates of equations (3) and (4) for ticket bookings departing within a day and 2–7 days. For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(2) and (3)–(4) report the results for train and airline ticket bookings, respectively. Columns (1) and (3) show the results for intra-province tickets and columns (2) and (4) report the results for inter-province tickets. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. The Fisher's permutation test is used to test the statistical significance of the difference between two groups of the estimator. The *p*-values are calculated using bootstrapping procedure.

Table A.9
Heterogeneity – 4A + Attractions (Train).

	(1)	(2)	(3)	(4)	(5)	(6)
	Intra-province			Inter-province		
Panel A: 0–1 days						
AQI difference	0.084	0.284	1.265***	0.006***	0.008***	0.014***
	(0.062)	(0.2)	(0.351)	(0.002)	(0.003)	(0.006)
Observations	892,589	1,140,858	1,156,300	3,738,233	5,559,829	9,786,149
Benchmark	270.563	535.905	1074.528	34.158	41.545	89.304
Relative effect (%)	1.556	2.652	5.887	0.898	0.919	0.794
KP F stat.	60.44	137.4	67.83	249	214.7	161.8
Panel B: 2–7 days						
AQI difference	0.039	0.119	1.132	0.009*	0.005	0.029***
	(0.093)	(0.33)	(0.74)	(0.005)	(0.004)	(0.011)
Observations	745,434	990,707	1,056,906	2,810,227	4,440,552	8,542,504
Benchmark	223.372	478.174	980.527	43.873	65.530	144.729
Relative effect (%)	0.877	1.241	5.774	1.067	0.392	0.998
KP F stat.	55.40	128.9	64.94	229.1	183	167.7
# of 4 A + attractions (xtile)	1	2	3	1	2	3
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table further presents heterogeneity in destination cities for train tickets. Panel A and B report the 2SLS estimates of equations (3) and (4) for ticket bookings departing within 0–1 days and 2–7 days. For each column, the dependent variable is the probability of individuals from an origin city *i* choosing to travel to destination city *j* on reservation date *t* in a specific booking period. Columns (1)–(3) show the results for intra-province tickets and columns (2)–(4) report the results for inter-province tickets. We divide the sample into three groups (xtiles) according to the number of 4 A-plus tourist attractions—a larger xtile means more 4 A-plus tourist attractions in the destination city. All regressions control linear and quadratic forms of weather difference between origin and destination city, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. *Benchmark* is the average of the dependent variable, and the *relative effect (%)* is the ratio of the estimated coefficient and *benchmark*, multiplied by 50. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A.10
The Role of Pollution Forecast on Travel Flows.

	(1)	(2)	(3)	(4)	(5)	(6)
	trips leaving in 0–1 days (using 24 h forecast)			trips leaving in 2–7 days (using 72 h forecast)		
Panel A: 2SLS						
AQI difference	−0.303	−0.306	−0.299	0.081	0.081	0.081
	(0.261)	(0.260)	(0.260)	(0.140)	(0.140)	(0.140)
AQI_origin_fcst_mean	0.524*			0.182*		
	(0.274)			(0.100)		
AQI_destin_fcst_mean	−0.085			0.058		
	(0.192)			(0.208)		
AQI_origin_fcst_min		0.538*			0.194*	
		(0.275)			(0.102)	
AQI_destin_fcst_min		−0.090			0.041	
		(0.194)			(0.216)	
AQI_origin_fcst_max			0.506*			0.170*
			(0.269)			(0.099)
AQI_destin_fcst_max			−0.077			0.073
			(0.187)			(0.200)
KP F stat.	226.4	222.7	229.7	256.5	256.3	256.5
Panel B: OLS						
AQI difference	0.061	0.060	0.063	0.149**	0.148*	0.149**
	(0.055)	(0.052)	(0.057)	(0.074)	(0.074)	(0.074)
AQI_origin_fcst_mean	0.280**			0.178*		
	(0.136)			(0.098)		
AQI_destin_fcst_mean	0.160			0.063		
	(0.121)			(0.205)		
AQI_origin_fcst_min		0.289**			0.189*	
		(0.135)			(0.099)	
AQI_destin_fcst_min		0.160			0.046	
		(0.124)			(0.213)	
AQI_origin_fcst_max			0.269*			0.166*
			(0.136)			(0.097)
AQI_destin_fcst_max			0.159			0.078
			(0.118)			(0.196)
Observations	252,772	252,772	252,772	252,275	252,275	252,275
CityPair FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Destin-YearMonth FE	Yes	Yes	Yes	Yes	Yes	Yes
Origin-YearMonth FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents the role of pollution forecasts using a small sample of data collected over five months, from September 2023 to January 2024, for 43 cities. Column (1)–(3) present the results for trips leaving in 0–1 days, and columns (4)–(6) presents the results for trips leaving in 2–7 days. For travels departing within a 0–1 day window, we utilize the 24-h AQI forecast, and for travels departing in 2–7 days, we use the 72-h AQI forecast. All regressions control for linear and quadratic forms of weather differences between origin and destination cities, consisting of temperature, wind speed, precipitation, sunshine, relative humidity, and atmospheric pressure. Standard errors in parentheses are robustly two-way clustered at the origin- and destination-city level.

References

Barwick, Panle Jia, Donaldson, Dave, Li, Shanjun, Lin, Yatang, Rao, Deyu, 2021. Improved Transportation Infrastructure Facilitates Adaptation to Pollution and Temperature Extremes. Available at: SSRN 3814060.

Barwick, Panle Jia, Li, Shanjun, Rao, Deyu, Zahur, Nahim Bin, 2024. The healthcare cost of air pollution: evidence from the world's largest payment network. Rev. Econ. Stat. https://doi.org/10.1162/rest_a_01430.

Bayer, Patrick, Keohane, Nathaniel, Timmins, Christopher, 2009. Migration and hedonic valuation: the case of air quality. J. Environ. Econ. Manag. 58 (1), 1–14.

Busse, Meghan R., Pope, Devin G., Pope, Jaren C., Silva-Risso, Jorge, 2015. The psychological effect of weather on car purchases. Q. J. Econ. 130 (1), 371–414.

Chang, Tom Y., Zivin, Joshua Graff, Gross, Tal, Neidell, Matthew, 2019. The effect of pollution on worker productivity: evidence from call center workers in China. Am. Econ. J. Appl. Econ. 11 (1), 151–172.

Chang, Tom Y., Huang, Wei, Wang, Yongxiang, 2018. Something in the air: pollution and the demand for health insurance. Rev. Econ. Stud. 85 (3), 1609–1634.

Chen, Shuai, Chen, Yuyu, Lei, Ziteng, Tan-Soo, Jie-Sheng, 2021. Chasing clean air: pollution-induced travels in China. J. Assoc. Environ. Resour. Econ. 8 (1), 59–89.

Chen, Shuai, Chen, Yuyu, Lei, Ziteng, Tan-Soo, Jie-Sheng, 2020. Impact of air pollution on short-term movements: evidence from air travels in China. J. Econ. Geogr. 20 (4), 939–968.

Chen, Siyu, Guo, Chongshan, Huang, Xinfei, 2018. Air pollution, student health, and School absences: evidence from China. J. Environ. Econ. Manag. 92, 465–497.

Chen, Yuyu, Ebenstein, Avraham, Greenstone, Michael, Li, Hongbin, 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 (32), 12936–12941.

Chew, Soo Hong, Huang, Wei, Li, Xun, 2021. Does haze cloud decision making? A natural laboratory experiment. J. Econ. Behav. Organ. 182, 132–161.

Conlin, Michael, O'Donoghue, Ted, Vogelsang, Timothy J., 2007. Projection bias in catalog orders. Am. Econ. Rev. 97 (4), 1217–1249.

Currie, Janet, Neidell, Matthew, 2005. Air pollution and infant health: what can we learn from California's recent experience? Q. J. Econ. 120 (3), 1003–1030.

Ebenstein, Avraham, Lavy, Victor, Roth, Sefi, 2016. The long-run economic consequences of high-stakes examinations: evidence from transitory variation in pollution. Am. Econ. J. Appl. Econ. 8 (4), 36–65.

Fan, Maoyong, He, Guojun, Zhou, Maigeng, 2020. The winter choke: coal-fired heating, air pollution, and mortality in China. J. Health Econ. 71, 102316.

Fang, Hanming, Wang, Long, Yang, Yang, 2020. Competition and Quality: Evidence from High-Speed Railways and Airlines. NBER working paper 27475.

Freeman, Richard, Liang, Wenquan, Song, Ran, Timmins, Christopher, 2019. Willingness to pay for clean air in China. J. Environ. Econ. Manag. 94, 188–216.

Guarnieri, Michael, Balmes, John R., 2014. Outdoor air pollution and asthma. Lancet 383 (9928), 1581–1592.

He, Jiaxiu, Liu, Haoming, Salvo, Alberto, 2019. Severe air pollution and labor productivity: evidence from industrial towns in China. Am. Econ. J. Appl. Econ. 11 (1), 173–201.

Herrnstadt, Evan, Heyes, Anthony, Muehlegger, Erich, Saberian, Soodeh, 2021. Air pollution and criminal activity: microgeographic evidence from chicago. Am. Econ. J. Appl. Econ. 13 (4), 70–100.
Ito, Koichiro, Zhang, Shuang, 2020. Willingness to pay for clean air: evidence from air purifier markets in China. J. Polit. Econ. 128 (5), 1627–1672.
Jia, Ruixue, Ku, Hyejin, 2019. Is China's pollution the culprit for the choking of South Korea? Evidence from the asian dust. Econ. J. 129 (624), 3154–3188.
Jung, Miyeon, Cho, Daegon, Shin, Kwangsoo, 2019. The impact of particulate matter on outdoor activity and mental health: a matching approach. Int. J. Environ. Res. Publ. Health 16 (16), 2983.
Khanna, Gaurav, Liang, Wenquan, Mobarak, Ahmed Mushfiq, Song, Ran, 2021. The Productivity Consequences of Pollution-Induced Migration in China. NBER working paper 28401.
Knittel, Christopher R., Douglas, L Miller, Sanders, Nicholas J., 2016. Caution, drivers! Children present: traffic, pollution, and infant health. Rev. Econ. Stat. 98 (2), 350–366.
Landrigan, Philip J., Fuller, Richard, Acosta, Nereus JR., Adeyi, Olusoji, Arnold, Robert, Baldé, Abdoulaye Bibi, Bertollini, Roberto, Bose-O'Reilly, Stephan, Boufford, Jo Ivey, Breysse, Patrick N., 2018. The lancet commission on pollution and health. Lancet 391 (10119), 462–512.
Li, Jennifer Jie, Massa, Massimo, Zhang, Hong, Zhang, Jian, 2019. Air pollution, behavioral bias, and the disposition effect in China. J. Financ. Econ. 142 (2), 641–673.
Lin, Yatang, Qin, Yu, Wu, Jing, Xu, Mandi, 2021. Impact of high-speed rail on road traffic and greenhouse gas emissions. Nat. Clim. Change 11, 952–957.
Qin, Yu, Wu, Jing, Yan, Jubo, 2019. Negotiating housing deal on a polluted day: consequences and possible explanations. J. Environ. Econ. Manag. 94, 161–187.
Qin, Yu, Zhu, Hongjia, 2018. Run away? Air pollution and emigration interests in China. J. Popul. Econ. 31 (1), 235–266.
Sanderson, Eleanor, Windmeijer, Frank, 2016. A weak instrument F-test in linear iv models with multiple endogenous variables. J. Econom. 190 (2), 212–221.
Sun, Cong, Zheng, Siqi, Wang, Jianghao, Kahn, Matthew E., 2019. Does clean air increase the demand for the consumer city? Evidence from beijing. J. Reg. Sci. 59 (3), 409–434.
Barwick, Panle Jia, Li, Shanjun, Lin, Liguo, Zou, Eric Yongchen, 2024. From Fog to Smog: The Value of Pollution Information. Am. Econ. Rev. 114 (5), 1338–1381.
Triguero-Mas, Margarita, Donaire-Gonzalez, David, Seto, Edmund, Valentín, Antònia, Martínez, David, Smith, Graham, Hurst, Gemma, Carrasco-Turigas, Glòria, Masterson, Daniel, Berg, Magdalena van den, Ambròs, Albert, Martínez-Íñiguez, Tania, Dedele, Audrius, Ellis, Naomi, Grazulevicius, Tomas, Voorsmit, Martin, Cirach, Marta, Cirac-Claveras, Judith, Swart, Wim, Clasquin, Eddy, Ruijsbroek, Annemarie, Maas, Jolanda, Jerret, Michael, Gražulevičienė, Regina, Kruize, Hanneke, Gidlow, Christopher.J., Nieuwenhuijsen, Mark.J., 2017. Natural outdoor environments and mental health: stress as a possible mechanism. Environ. Res. 159, 629–638. https://www.sciencedirect.com/science/article/abs/pii/S0013935117307375?via%3Dihub.
Wang, Jianjun, Zhang, Meigen, Bai, Xiaolin, Tan, Hongjian, Li, Sabrina, Liu, Jiping, Zhang, Rui, Wolters, Mark A., Qin, Xiuyuan, Zhang, Miming, 2017. Large-scale transport of Pm 2.5 in the lower troposphere during winter cold surges in China. Sci. Rep. 7 (1), 1–10.
Williams, Austin M., 2019. Understanding the micro-determinants of defensive behaviors against pollution. Ecol. Econ. 163, 42–51.
Zhang, Anming, Wan, Yulai, Yang, Hangjun, 2019. Impacts of high-speed rail on airlines, airports and regional economies: a survey of recent research. Transport Pol. 81, A1–A19.
Zhang, Junjie, Mu, Quan, 2018. Air pollution and defensive expenditures: evidence from particulate-filtering facemasks. J. Environ. Econ. Manag. 92, 517–536.
Zhang, Lin, Liu, Licheng, Zhao, Yuanhong, Gong, Sunling, Zhang, Xiaoye, Henze, Daven K., Capps, Shannon L., Fu, Tzung-May, Zhang, Qiang, Wang, Yuxuan, 2015. Source attribution of particulate matter pollution over north China with the adjoint method. Environ. Res. Lett. 10 (8), 084011.
Zhang, Xin, Chen, Xi, Zhang, Xiaobo, 2018. The impact of exposure to air pollution on cognitive performance. Proc. Natl. Acad. Sci. USA 115 (37), 9193–9197.
Zhang, Xin, Zhang, Xiaobo, Chen, Xi, 2017. Happiness in the air: how does a dirty sky affect mental health and subjective well-being? J. Environ. Econ. Manag. 85, 81–94.