Chasing Clean Air

> Chen, Shuai, et al. "Chasing Clean Air: Pollution-Induced Travels in China." *Journal of the Association of Environmental and Resource Economists* 8.1 (2021): 59-89.

Authors

Shuai Chen

- Associate Professor, China Academy for Rural Development and the School of Public Affairs, Zhejiang University
- **Research Interest**: climate change, rural labor supply, air pollution, and human migration
- **Publication**: JEEM(2), …

Yuyu Chen

- Professor, Guanghua School of Management, Peking University
- **Research Interest**: Ineqaulity, Human capital, health, Productivity and growth
- **Publication**: AER(2), JPE, PNAS, REStat, …

Ziteng Lei

- PhD Candidate, Department of Economics, UC Santa Barbara
- **Research Interest**: labor economics, environmental economics, and applied econometrics
- **Publication**: JEG, …

Jie-Sheng Tan-Soo

- Assistant Professor, Lee Kuan Yew School of Public Policy, National University of Singapore
- **Research Interest**: Ecosystem services valuation, Global environmental health and development, Pollution regulation in China
- **Publication**: PNAS, Energy Policy, JEG, …

1 Background

1.1 How literature tackles air pollution issues

- The first way is to **estimate the direct and indirect health impacts of air pollution**.
- The second way is to **derive a willingness-to-pay (WTP) measure for air quality improvements** using stated or actual behavioral responses.
    - Conduct contingent valuation method (CVM) study
    - OR rely on observed behaviors to value air quality

> **One central difference** between these two strategies is that the latter relies on individuals actively responding to their surrounding air quality.

1.2 Averting behaviors

- **Migration:** Zheng and Kahn (2008); Bayer et al. (2009); S. Chen et al. (2017); Qin and Zhu (2018); Tan-Soo (2018); Freeman et al. (2019)
- **Purchases of face masks & air purifiers, and health expenditures:** Cong Sun et al. (2017); Deschênes et al. (2017); Barwick et al. (2018); Zhang and Mu (2018); Ito and Zhang (2020)

1.3 Permanent migration v.s. short-run travel

- Patterns of **permanent migration** between locations are induced by their differences in air quality, among other factors (e.g., Sieg et al. 2004; Bayer et al. 2009; S. Chen et al. 2017; Tan-Soo 2018; Freeman et al. 2019).
- On the other hand, **short-term movements** on "bad days" can also be an appropriate travel-based avoidance response (Neidell, 2009; Graff Zivin and Neidell, 2009).
    - But only self-reported evidence from China.

1.4 What this paper does

Idea: Using mobile phones' signal data from Telecom, this paper investigates the relationship between air quality differences and population movement between pairs of Chinese cities on any given day.

Hypothesis: At the daily level, there is a net flow of people moving from cities with poorer air quality toward cities with better air quality.

Empirical challenge: more pollution relates to more economic opportunities. $\Rightarrow$ underestimation

Identification strategy: long-range transmission of upwind air pollution as IV

Further Work: detect nonlinear relationships and decision-making mechanisms

1.5 Literature Contribution

1. This is one of the first studies to investigate how short-term movement is induced by locational differences in public goods.
2. Using short-term movement to value air quality confers an empirical advantage over using permanent migration as the former is motivated by a simpler decision-making process.
3. We take care to establish the causal relationship between air pollution and population flows.
4. We demonstrate empirically that all else being equal, places with clean air attract more visitors relative to places with more polluted air.

2 Model & Empirical Strategy

2.1 Theorical Motivation

$$U = U(X, L, S) \tag{1}$$

where

$$S = S(P, Z). \tag{2}$$

- X: consumption of a unitaire good
- L: leisure time
- S: time spent sick
- P: ambient air pollution
- Z: other exogenous factors

Budget constraint:

$$I + wT_w = X + wL + c + M(S) + wS. \tag{3}$$

- I: nonlabor income
- w: wage rate
- T_w: total time endowment
- c: net cost of traveling (positive or negative)
- $M(S)$: medical expenses

The averting strategy here is a choice of location:

- $A = 0$: staying in city o, $P = P_o$, $c = 0$
- $A = 1$: travel to location d, $P = P_d$, $c = c_d$

$$\max_{A,X,L} U = U(X, L, S)$$

$$s.t. \quad I + wT_w = X + wL + c + M(S) + wS; \tag{4}$$

$$S = S(P, Z); A \in \{0, 1\}; P = P_o, c = 0 \; if \;\; A = 0;$$
$$P = P_d, c = c_d \; if \;\; A = 1.$$

> Solve it in two steps:
>
> - First, independent of A, the individual maximizes utility by choosing consumption X and leisure L. $\Rightarrow V(I, w, S(P), c)$
> - Second, the individual now chooses A (travel or not) to maximize utility. Choose to travel if $V(I, w, S(P_o), 0) \leq V(I, w, S(P_d), c_d)$, and vice versa.

Suppose that the net travel cost of individual c_d is varied for different individuals and follows a distribution with differentiable cumulative density function $F(c)$ with support $[\underline{c}, \bar{c}]$. By assuming

- (i) the probability of an individual in city o willing to travel to city d to be $Pr(Travel_{od})$,
- (ii) the population in city o is N_o,
- (iii) the population in city d is N_d,
  the net-flow ratio for route od is defined as:

$$NetFlowRatio_{od} = \frac{Flow_{od} - Flow_{do}}{Flow_{od} + Flow_{do}} = \frac{N_o \times Pr(Travel_{od}) - N_d \times Pr(Travel_{do})}{N_o \times Pr(Travel_{od}) + N_d \times Pr(Travel_{do})} \tag{5}.$$

We hypothesize that $\partial NF_{od}/\partial P_o \geq 0, \partial NF_{od}/\partial P_o \leq 0$

2.2 Empirical Strategy

$$NF_{ijt} = \beta_0 + \beta_1(P_{it} - P_{jt}) + (W_{it} - W_{jt})\theta + D_t + \gamma_{jt} + \phi_{ij} + \varepsilon_{ijt}, \tag{6}$$

where $NF_{ijt} = (F_{ijt} - F_{jit})/(F_{ijt} + F_{jit})$.

2.3 Instrumental Variable

It is likely that P is positively correlated with economic activity (E), which in turn is a pull factor for incoming travelers. $\Rightarrow \beta_1$ will likely be underestimated using OLS estimation.

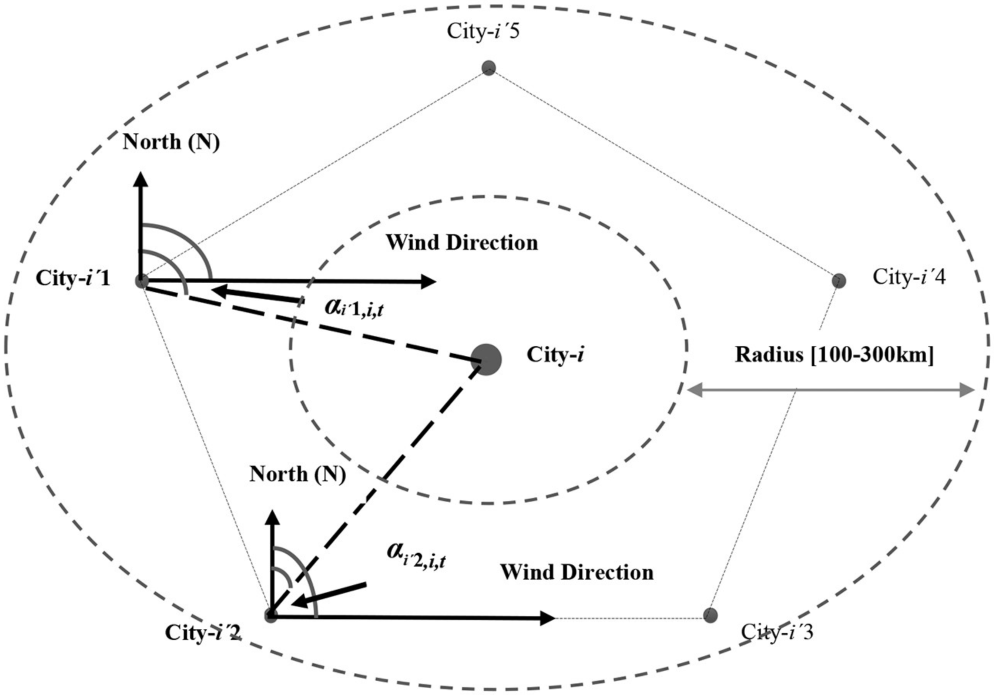
Toward this end, we construct an IV based on **wind direction** and **long-range upwind transmission of air pollutants** (e.g., Bayer et al. 2009; Barwick et al. 2018; Tan-Soo 2018; Deryugina et al. 2019).

Following Barwick et al. (2018),

$$UP_{it} = \sum_{i' \neq i} \frac{P_{i't} \times WS_{i't}}{D_{i'i}^2} \times cos\theta_{i',i,t} \tag{7}$$

where

$$cos\theta_{i',i,t} = \max\{cos(\gamma_{i't} - \beta_{i't}), 0\} \tag{8}$$



For the exclusion restriction condition to be violated, the confounding variable(s) would need to vary at similar pulses as upwind locations'

- (i) distance to downwind location,
- (ii) daily air quality,
- (iii) daily wind direction,
- (iv) daily wind speed.

In this regard, the exclusion restriction condition may be violated in two ways:

- First, the upwind locations may share many similar characteristics (other than air quality) if they are too near to the downwind location for which they are instrumenting.
- Second, even though the set of cities instrumenting for say, city A is weighted by distance, daily wind direction, speed, and air quality, it is still plausible that the set of instrumenting cities affect travels between city A and city B by being alternate travel destinations.

$$NF_{ijt} = \alpha_0 + \alpha_1\widehat{(P_{it} - P_{jt})} + (W_{it} - W_{jt})\theta + D_t + \varphi_{ij} + \varepsilon_{ijt} \tag{6'}$$

where

$$(P_{it} - P_{jt}) = \gamma_0 + \gamma_1(UP_{it} - UP_{jt}) + (W_{it} - W_{jt})\theta + D_t + \varphi_{ij} + \mu_{ijt} \tag{9}$$

3 Data

3.1 Data Sources

First, the raw data set for calculating population movement between cities is from China Telecom. China Telecom provided data of population flows between a group of cities (25 cities in March 2016 and three cities in June 2016), recording all flows of China Telecom mobile users between these given cities on a 24-hour basis.

Table 1. Individual Observations and City Sample

			Observations for March 2016		Observations for June 2016	
City Name	Province	City Status	Outflows	Inflows	Outflows	Inflows
Beijing	Beijing	Provincial city	596,375	557,052	243,318	248,336
Tianjin	Tianjin	Provincial city	532,247	577,573	NA	NA
Chengde	Hebei	Prefecture city	34,019	44,546	NA	NA
Hengshui	Hebei	Prefecture city	49,761	70,862	NA	NA
Shanghai	Shanghai	Provincial city	625,224	770,848	215,080	210,057
Nanjing	Jiangsu	Provincial capital	320,297	304,663	NA	NA
Wuxi	Jiangsu	Prefecture city	371,804	362,557	NA	NA
Suzhou	Jiangsu	Prefecture city	852,148	753,529	NA	NA
Hangzhou	Zhejiang	Provincial capital	534,886	511,022	NA	NA
Ningbo	Zhejiang	Prefecture city	191,138	180,435	NA	NA
Lishui	Zhejiang	Prefecture city	16,738	17,174	NA	NA
Xuancheng	Anhui	Prefecture city	40,653	70,407	NA	NA
Wuhan	Hubei	Provincial capital	25,289	28,201	NA	NA
Xiangyang	Hubei	Prefecture city	24,809	21,123	NA	NA
Xianning	Hubei	Prefecture city	61,729	62,030	NA	NA
Guangzhou	Guangdong	Provincial capital	862,059	908,008	NA	NA
Shenzhen	Guangdong	Subprovincial city	446,926	473,968	196,536	196,541
Heyuan	Guangdong	Prefecture city	133,969	126,967	NA	NA
Yunfu	Guangdong	Prefecture city	113,490	110,920	NA	NA
Chongqing	Chongqing	Provincial city	403,718	366,915	NA	NA
Chengdu	Sichuan	Provincial capital	583,925	591,135	NA	NA
Mianyang	Sichuan	Prefecture city	175,044	151,579	NA	NA
Ya'An	Sichuan	Prefecture city	128,991	130,639	NA	NA
Bazhong	Sichuan	Prefecture city	35,837	25,431	NA	NA
Zunyi	Guizhou	Prefecture city	134,739	78,231	NA	NA
Total	12 provinces	25 cities	7,295,815	7,295,815	654,934	654,934

Note. Individual sample (total observation for individual movement) = 7,950,749. City sample consists of 25 cities from 12 province, covering four administrative levels. Bidirectional population flow routes = 600, with 390 nonzero flow routes.

The second source of data pertains to air quality. Air pollution data are obtained from the website of the China National Environmental Monitoring Center (CNEMC), which is affiliated with the Ministry of Environmental Protection of China (MEP).

Third, weather and climate data are obtained from the China Meteorological Data Service Center (CMDC), which is affiliated with the National Meteorological Information Center of China.

We match air pollution data and meteorological conditions from stations to Chinese cities using the inverse-distance weighting (IDW) method.

3.2 Descriptive Statistics

Table 2. Descriptive Statistics

		Bidirectional (N = 11,823)				Unidirectional-1 (N = 5,911)		Unidirectional-2 (N = 5,912)	
Variable	Definition/Unit	Mean	SD	Min	Max	Mean	SD	Mean	SD
Population flows:									
Outflows	Number of people	672.5	2,057	0	72,500	653.0	1,988	692.0	2,123
Outflow percentage	Outflow/(outflows + inflows) × 100%	50.01	16.57	0	100	47.05	16.31	52.97	16.30
Inflows	Number of people	672.5	2,057	0	72,500	692.1	2,123	652.9	1,988
Inflow percentage	Inflow/(outflows + inflows) × 100%	50.01	16.56	0	100	52.97	16.28	47.05	16.30
Netflow	Number of people	0	872.8	−25700	25,700	−39.11	871.3	39.11	872.5
Netflow percentage	(Outflows − inflows)/(outflows + inflows) × 100%	.021	33.15	−100	100	−5.892	32.63	5.933	32.60
Air pollution:									
AQI	Air quality index (0–500)	87.89	46.48	14.00	344.0	96.37	54.27	79.40	35.09
AQI difference	AQI(origin) − AQI(destination)	.002	54.72	−294.8	294.8	16.97	52.03	−16.96	52.02
Abs(AQI difference)	Absolute value of AQI difference	37.53	39.83	.002	294.8	37.53	39.83	37.52	39.82

Note. Unidirectional route = 195; bidirectional route = 390. Descriptive statistics for other controls in table A4.

4 Results

4.1 Baseline Results

Table 3. Main Regression Results

	(1)	(2)	(3)	(4)
	Dependent Variable: AQI Difference			
A. 2SLS:				
First-stage estimation:				
Upwind AQI difference	27.9576***	38.2303***	37.7937***	34.9409***
	(2.5045)	(2.5559)	(2.5483)	(2.4519)
First-stage *F*-statistics	124.6	162.0	146.7	156.0
	Dependent Variable: Net-Flow Percentage			
Second-stage estimation:				
AQI difference	.1177**	.1331***	.1333***	.1460***
	(.0461)	(.0346)	(.0353)	(.0412)
KP F-statistics	124.6	223.7	220.0	203.1
	Dependent Variable: Net-Flow Percentage			
B. OLS:				
AQI difference	.0129**	.0181***	.0179**	.0192***
	(.0064)	(.0070)	(.0069)	(.0074)
R-squared	.2081	.2120	.2120	.2334
Route FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Weather controls	No	Yes	Yes	Yes
Distance-weighted average AQI of surrounding cities	No	No	Yes	Yes
Amenities	No	No	No	Yes

Note. $N = 11{,}823$. Date fixed effects (FE) include month fixed effects, day-of-week fixed effects, and holiday fixed effects. Weather controls are included as the difference between two cities and consist of minimum, average, and maximum temperature, precipitation, hours of daylight, relative humidity, and wind speed. Regressions are weighted by the market share of China Telecom. Standard errors are listed in parentheses and clustered on 390 flow routes.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

4.2 Origin versus Destination AQI

Table 4. Investigation of Origin versus Destination Air Quality in Affecting Movements

	Dependent Variable			
	Net-Flow Percentage		Outflow Percentage	Inflow Percentage
	(1)	(2)	(3)	(4)
AQI difference (baseline)	.1460***		.0730***	−.0718***
	(.0412)		(.0206)	(.0206)
AQI origin based		.1573***		
		(.0438)		
AQI destination based		−.1333***		
		(.0450)		
Route FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Distance-weighted average AQI of surrounding cities	Yes	Yes	Yes	Yes
Amenities	Yes	Yes	Yes	Yes
KP F-statistics	203.1	101.8	203.1	203.1

Note. N = 11,823. In col. 2, we use upwind pollution transmission both to the origin city and to the destination city as IVs for the two endogenous air pollution variables. Date FE includes month fixed effects, day-of-week fixed effects, and holiday fixed effects. Weather controls are included as the difference between two cities and consist of minimum, average, and maximum temperature, precipitation, hours of daylight, relative humidity, and wind speed. Regressions are weighted by the market share of China Telecom. Standard errors are listed in parentheses and clustered on 390 flow routes. AQI = air quality index.

* $p < .1$.
** $p < .05$.
*** $p < .01$.

4.3 Robustness Checks

Table 5. Various Robustness Checks of the Regression Results

	Dependent Variable: Net-Flow Percentage							
	Standard Errors			IV Radius (km)			Unidirectional Routes	
	Bootstrap	Two Way	HAC-Auto	100–200	150–300	300–500	O-D	D-O
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AQI difference	.1460***	.1460**	.1460**	.0923***	.1553***	.5738	.1539***	.1265**
	(.0485)	(.0641)	(.0636)	(.0324)	(.0393)	(.6462)	(.0555)	(.0625)
Observations	11,823	11,823	11,823	11,823	11,823	11,823	5,911	5,912
No. of flow routes	390	390	390	390	390	390	195	195
Cluster	City-boots	Route-date	500 km-(2)	Route	Route	Route	Route	Route
IV range (km)	100–300	100–300	100–300	100–200	150–300	300–500	100–300	100–300
KP F-statistics	41.64	37.97	. . .	302.8	266.9	4.296	107.8	92.99

Note. In line with the baseline model, all regression models control for route and date fixed effects, weather conditions, distance-weighted average air quality index (AQI) of surrounding cities, and amenities, and are weighted by the market share of China Telecom. In col. 3, we estimate standard errors that allow for spatial correlation within 500 km and temporal correlation within two time periods (results from alternative selections of spatial and temporal correlations are reported in table A9). In addition to cols. 4–6, we also report the robustness of other ranges of IV radius in table A10. Unless otherwise noted, standard errors are listed in parentheses and clustered on 390 flow routes. O = origin; D = destination.

* $p < .1$.
** $p < .05$.
*** $p < .01$.

4.4 Nonlinear Responses

4.4.1 Distance and Magnitude of AQI Differences between Cities

Table 6. Regression Results by Distance and by AQI Difference between Cities

	Dependent Variable: Net-Flow Percentage			
	(1)	(2)	(3)	(4)
	A. Distance by AQI Difference			
AQI difference	.7376**	.3388**	.1199	–.0259
	(.3095)	(.1548)	(.0866)	(.0462)
N	3,463	2,344	2,436	3,580
AQI difference	>=50	>=50	<50	<50
Distance (km)	<1,000	>=1,000	<1,000	>=1,000
KP F-statistics	19.85	31.17	46.42	106.3
	B. By Either Distance or AQI Difference			
AQI difference	.4373***	.0380	.2611***	.0746**
	(.1137)	(.0387)	(.0901)	(.0321)
N	5,807	6,016	5,899	5,924
AQI difference	>=50	<50	Total	Total
Distance (km)	Total	Total	<1,000	>=1,000
KP F-statistics	45.49	178.3	55.15	168.3

Note. In line with the baseline model, all regression models control for route and date fixed effects, weather conditions, distance-weighted average air quality index (AQI) of surrounding cities, and amenities, and are weighted by the market share of China Telecom. Standard errors are listed in parentheses and clustered on flow routes.

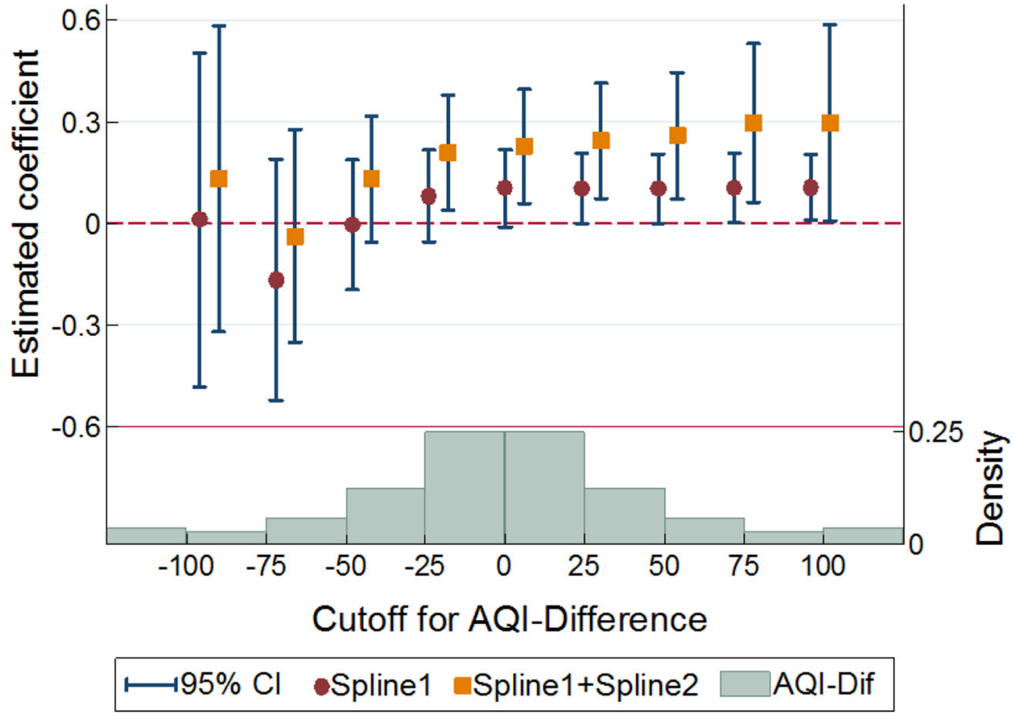
* $p < .1$.

** $p < .05$.

*** $p < .01$.

4.4.2 Spline Regression

$$NF_{ijt} = \beta_0 + \beta_1(\widehat{P_{it} - P_{jt}}) + \beta_2\left[(\widehat{P_{it} - P_{jt}}) - \kappa\right] \cdot I((P_{it} - P_{jt}) \geq \kappa)$$
$$+ (W_{it} - W_{jt})\theta + D_t + \gamma_{jt} + \varphi_{ij} + \varepsilon_{ijt} \qquad (10)$$



4.5 Evidence on the Travel Decision-Making Process

4.5.1 Duration of Stay

Table 7. Effects of Air Pollution on Net-Flow Ratio: By Travel Duration

	Dependent Variable: Net-Flow Percentage									
	Duration of Stay									
	Total (1)	<=1 Day (2)	<=2 Days (3)	<=3 Days (4)	<=4 Days (5)	<=5 Days (6)	<=6 Days (7)	2–3 Days (8)	4–5 Days (9)	≥6 Days (10)
AQI difference	.1460***	.0163	.1488***	.1770***	.1663***	.1393***	.1375***	.1912***	.1157**	.0356
	(.0412)	(.0589)	(.0479)	(.0460)	(.0429)	(.0437)	(.0436)	(.0508)	(.0534)	(.0536)
Duration quantile	0–100	0–22.4%	0–61.0%	0–70.8%	0–77.0%	0–81.5%	0–84.9%	22.4–70.8%	70.8–81.5%	84.88–100%
Observations	11,823	10,326	11,427	11,556	11,629	11,677	11,709	10,509	11,501	10,242
KP *F*-statistics	203.1	188.9	209	206.3	205.6	201.7	202.4	187.7	208.7	192.0

Note. In line with the baseline model, all regression models control for route and date fixed effects, weather conditions, distance-weighted average air quality index (AQI) of surrounding cities, and amenities, and are weighted by the market share of China Telecom. Standard errors are listed in parentheses and clustered on flow routes.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

4.5.2 Lead Air Quality at Destination

$$NF_{ijt} = \alpha_0 + \alpha_1 \widehat{(P_{it} - Lead_{j,t+k})} + (W_{it} - W_{jt})\theta + D_t + \varphi_{ij} + \varepsilon_{ijt} \quad (11)$$

where

$$(P_{it} - Lead_{j,t+k}) = \gamma_0 + \gamma_1 (UP_{it} - UP_{jt}) + (W_{it} - W_{jt})\theta + D_t + \varphi_{ij} + \mu_{ijt} \quad (12)$$

4.5.3 Trips on Workdays v.s. Nonworkdays

Table 8. Working Periods versus Nonworking Periods

	Dependent Variable: Net-Flow Percentage				
	Total (1)	Trips outside Working Hours (2)	Trips during Working Hours (3)	Trips during Nonworkdays (4)	Trips during Workdays (5)
Age = (21–60]					
AQI difference	.1783***	.1945***	.0464	.3938***	.1546***
	(.0423)	(.0447)	(.0534)	(.0884)	(.0357)
Observations	11,781	11,594	8,578	4,805	9,601
KP F-statistics	206.6	206.6	187.1	69.39	207.8
Percent share in obs.	93.3%	46.2%	47.1%	21.5%	71.7%

Note. In line with the baseline model, all regression models control for route and date fixed effects, weather conditions, distance-weighted average air quality index (AQI) of surrounding cities, and amenities, and are weighted by the market share of China Telecom. Standard errors are listed in parentheses and clustered on flow routes.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

5 Conclusion

5.1 Summary

This study utilizes a novel data set of telecommunication signals from China to examine an effective but understudied averting behavior—short-term travel to avoid air pollution.

For every unit difference in AQI between a pair of cities, movement to less polluted city increases by around 0.15%, which means around 5.7% of travel.

Further examination about decision making of avoidance travel:

- First, nearer destinations are more attractive than further ones.
- Second and related, nonlinear relationship between air quality difference and movements.
- Third, trips made on nonworking days and outside of working hours exhibit a stronger relationship with the AQI difference.
- Fourth, most avoidance travels last for more than 1 day and less than 4–5 days. Average marginal willingness to pay (MWTP) for AQI is at around CNY 6.4.

5.2 Limitation

- First, information used in this study is limited temporally and spatially as it is only available for 2 months—March and June—and for 25 cities.

- Second, the empirical strategy employed in this study is limited by the type of observations available.
- Third, to the extent that travel is multipurpose, the effect we have estimated may be overstated.

5.3 Research and Policy Implications

- First, we have taken a first look at how Chinese residents use short-term travel to avoid air pollution.
- Second, Our results shed additional light on the decision-making processes behind avoidance travels, including the spatial and temporal circumstances where they are more likely to be observed.