NBER WORKING PAPER SERIES

TRANSPORTATION NETWORKS, SHORT-TERM MOBILITY, AND POLLUTION EXPOSURE: EVIDENCE FROM HIGH-SPEED RAIL IN CHINA

Panle Jia Barwick
Dave Donaldson
Shanjun Li
Yatang Lin
Deyu Rao

Working Paper 30462
http://www.nber.org/papers/w30462

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2022

We are grateful to Susanna Berkouwer, Fiona Burlig, Esther Duflo, Michael Greenstone, Paulina Oliva, Joseph Shapiro, Matthew Turner and Eric Zou for helpful comments, to Nahim Bin Zahur and Ziye Zhang for excellent research assistance, and to Minwei Tang for data support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Transportation Networks, Short-term Mobility, and Pollution Exposure: Evidence from High-Speed Rail in China
Panle Jia Barwick, Dave Donaldson, Shanjun Li, Yatang Lin, and Deyu Rao
NBER Working Paper No. 30462
September 2022
JEL No. O18,Q53,Q54,R41

ABSTRACT

Can the enhanced mobility created by transportation infrastructure investments help people to avoid environmental extremes? We use transaction records from China's card payment system to measure experienced pollution exposure (EPE)—that is, exposure based on the pollution levels at travelers' actual locations—and evaluate how EPE was affected by the country's high-speed railway network, even while holding pollution itself constant. Our estimates imply a reduction in EPE that corresponds to a mortality benefit of 21.3 million life-years saved, primarily due to travelers changing their destinations towards locations with predictably cleaner air.

Panle Jia Barwick
University of Wisconsin-Madison
Department of Economics
1180 Observatory Dr
Madison, WI 53706
and CEPR
and also NBER
panle.barwick@gmail.com

Dave Donaldson
Department of Economics, E52-554
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
ddonald@mit.edu

Shanjun Li
Cornell University
405 Warren Hall
Ithaca, NY 14853
and NBER
SL2448@cornell.edu

Yatang Lin
Department of Economics
Hong Kong University of Science and Technology
Hong Kong
yatang.lin@gmail.com

Deyu Rao
Department of Economics
Hong Kong University of Science and Technology
Hong Kong
dyrao@ust.hk

Transportation Networks, Short-term Mobility, and Pollution Exposure: Evidence from High-Speed Rail in China

PANLE JIA BARWICK DAVE DONALDSON SHANJUN LI
YATANG LIN DEYU RAO*

Can the enhanced mobility created by transportation infrastructure investments help people to avoid environmental extremes? We use transaction records from China's card payment system to measure experienced pollution exposure (EPE)—that is, exposure based on the pollution levels at travelers' actual locations—and evaluate how EPE was affected by the country's high-speed railway network, even while holding pollution itself constant. Our estimates imply a reduction in EPE that corresponds to a mortality benefit of 21.3 million life-years saved, primarily due to travelers changing their destinations towards locations with predictably cleaner air.

JEL: O18, Q53, Q54, R41

Keywords: Adaptation, Pollution, Mobility, Transportation Infrastructure

Environmental degradation and climate change present escalating risks, but ones that have an uneven spatial footprint.[1] Human mobility therefore offers a potential avenue for limiting exposure to environmental harm. While long-run migration has been the focus of widespread research, short-term travel—such as the "haze-avoidance tourism" described in (Arlt 2017) or (Sharkov 2016)—is a form of adaptation about which very little is known. This paper considers such mobility in the context of China, home to some of the world's most lethal air pollution. We first introduce new measures of *experienced pollution exposure* (EPE)—that is, the actual exposure of residents on the basis of their daily location choices.[2] We then use this concept to quantify whether the country's dramatic investments in transport infrastructure changed travel patterns in a way that reduced EPE even while holding constant the distribution of pollution itself.

Our analysis starts by creating a unique dataset of daily travel flows between all city-pairs in China, derived from in-person credit and debit card transactions within the world's largest interbank payment network (UnionPay). We then measure EPE by merging this dataset with high-frequency pollution readings throughout the country. Finally, we evaluate the causal effect of

* Barwick: Department of Economics, University of Wisconsin-Madison and NBER, pbarwick@wisc.edu; Donaldson: Department of Economics, Massachusetts Institute of Technology and NBER, ddonald@mit.edu; Li: Dyson School of Applied Economics and Management, Cornell University and NBER, SL2448@cornell.edu; Lin: Department of Economics, Hong Kong University of Science and Technology, linyt@ust.hk; Rao: Department of Economics, Hong Kong University of Science and Technology, dyrao@ust.hk. We are grateful to Susanna Berkouwer, Fiona Burlig, Esther Duflo, Michael Greenstone, Paulina Oliva, Joseph Shapiro, Matthew Turner and Eric Zou for helpful comments, to Nahim Bin Zahur and Ziye Zhang for excellent research assistance, and to Minwei Tang for data support.

[1] Existing work uses spatial research designs to document impacts of pollution and climate change on economic growth (e.g. Dell et al. 2012; Carleton and Hsiang 2016), social stability (e.g. Burke et al. 2010; Hsiang et al. 2013), and health (e.g. Deschênes et al. 2009; Currie and Neidell 2005; Chen et al. 2013; Landrigan et al. 2018; Deryugina et al. 2019; Carleton et al. 2022), among many other aspects.

[2] This is analogous to the concept of "experienced racial segregation" in Athey et al. (2021), which "captures individuals' exposure to diverse others in the places they visit over the course of their days" (p. 1).

the 2013-16 expansion of China's high-speed railway (HSR) system—whose 38,000 km length is more than twice as long as the combined length of all others in the world—on pollution exposure.

To do so, we develop a conceptual framework that decomposes the impact of HSR on national EPE, holding constant the nationwide distribution of daily air pollution, into effects that work on the extensive and intensive margins of travel responses. The extensive margin of EPE improves whenever HSR causes more inter-city travel and typical destinations offer better air quality than typical destinations. The intensive margin of EPE improves whenever HSR causes travelers to tilt their travel choices towards cleaner destinations. Neither of these effects has been studied before. Nor are the signs of these effects clear *a priori* given that highly polluted locations may nevertheless be attractive travel destinations for a host of other reasons.

We uncover two main findings. First, the effect of HSR expansion on the intensive margin of travel contributes to a substantial reduction in EPE: travelers' exposure to pollution drops by 14.7%. This implies that, every day, tens of millions of people enjoy an average reduction of ambient fine particulate matter ($PM_{2.5}$) by 2.5 μg/m^3 due to HSR. To put this in context, a reduction of this size translates (when using recent estimates of the mortality effects of pollution in China) into 21.3 million life-years saved. Second, we find that even though HSR causes a sizable expansion in the number of travelers, this extensive margin response actually increases EPE because, on average, travelers' destinations are slightly dirtier than their home cities. However, this effect is several orders of magnitude smaller than the intensive margin response, so it does little to change the fact that, on net, the mobility responses caused by China's HSR system did much to reduce EPE.

The final component of our analysis asks what types of adjustment can explain these effects. Perhaps surprisingly, we find that most of our total effect can be accounted for by travel to a cleaner set of equidistant destinations (rather than farther-flung ones), and to destinations that are predictably clean at times when travelers' origin cities are predictably dirty. That is, while the type of mobility we study involves short-term trips, its pollution-avoiding benefits derive because of travelers who respond to long-run averages rather than shorter-term deviations from such averages.

This study contributes to two strands of literature. The first examines adaptation to adverse environmental conditions (such as air pollution). Mechanisms under study have included reduced outdoor activity (e.g. Moretti and Neidell 2011; Graff Zivin and Neidell 2014; Barwick et al. 2024), increased defensive investments (e.g. Barreca et al. 2016; Zheng and Kahn 2017; Zhang and Mu 2018; Ito and Zhang 2020), altered longer-term migration (e.g. Banzhaf and Walsh 2008; Bayer et al. 2009; Deschênes and Moretti 2009; Kahn 2010; Black et al. 2011; Freeman et al. 2019; Balboni 2021; Chen et al. 2022; Cruz and Rossi-Hansberg 2024; Khanna et al. 2024), and limited work on shorter-term mobility (Chen et al. 2021). We add to this agenda by contributing the concept of EPE and its decomposition into intensive and extensive margin responses, the first comprehensive measurement of EPE, and an evaluation of how transport infrastructure investments cause changes in short-term mobility that can improve or exacerbate EPE.

Second, we contribute to the literature on quantifying the effects of transportation networks on, for example, enhanced mobility of people and goods (Baum-Snow 2010; Zheng and Kahn 2013; Faber 2014; Lin 2017; Donaldson 2018; Bernard et al. 2019), and changes in pollution emissions (Chang et al. 2021; Gendron-Carrier et al. 2022; Lee et al. 2023). In contrast, we study how such infrastructure investments can cause additional benefits by changing the composition and extent of passenger mobility in a way that improves aggregate pollution exposure, even holding pollution itself constant.

I. Data and Descriptive Evidence

Our analysis draws on a unique combination of data on population mobility, air pollution, and transportation infrastructure throughout China from 2013-2016. Summary statistics are presented in Table A1, with further details in Appendix A.

Intercity Travel Flow.—We construct daily bilateral passenger flows data from the universe of credit and debit card transactions conducted through the UnionPay network from January 2013 to December 2016.[3] Our sample tracks a randomly chosen 27 million cards (a 1% sample) via all of their transactions during 2013-2016. We focus exclusively on the 97% of transactions that are "offline" (i.e. where the cardholder is physically present at the merchant's location) to track the movement of card users domestically. We then map such locations to prefectures (henceforth "cities") and define any card's home city at date t as its most common location of use within a rolling 12-month window centered on t. Finally, we count a card from home city i as undergoing an inter-city trip to city j on date t whenever it is used in j on that date. The sum of all such trips, denoted $N_{ij,t}$, provides an estimate of the number of residents from city i who travel to city j on date t.[4]

Air Pollution Data.—Since 2013, China's Ministry of Ecology and Environment has used ground-level monitoring stations at over 1,000 sites and reported hourly air quality data on its website (Barwick et al. 2024).[5] We use these reports to calculate the daily concentration of fine particulate matter (measured as $PM_{2.5}$) at the city-day level (by averaging hourly readings from monitors within each city) and thereby construct various measures of pollution, denoted P_{it}. Our baseline analysis will use P_{it} defined as an indicator for the daily average $PM_{2.5}$ exceeding 75 μg/m^3, an official benchmark for extreme pollution. Across city-days from 2013-2016, the average $PM_{2.5}$ concentration was 52 μg/m^3 (SD = 44 μg/m^3) and 20% of observations exceeded 75 μg/m^3.[6]

Experienced Pollution Exposure.—At the heart of our analysis is a measure of the pollution that Chinese residents are exposed to, given their actual locations at any point in time. For any measure of pollution $P_{i,t}$ we define the *experienced pollution exposure* of city i on date t as

$$EPE_{i,t} = \sum_{j=1}^{J} N_{ij,t} \times P_{j,t},$$

which simply weights the pollution in each city $P_{j,t}$ by the number of residents of city i who are in

[3]As the sole interbank payment network in China, UnionPay tracked 34 trillion *yuan* (\$4.9 trillion) of purchases annually during the data period, which amounts to over 40% of retail consumption. Payment methods such as Wechat and Alipay were limited at the time, representing less than 2% of aggregate retail sales in 2013 and around 10% in 2016.

[4]Because this source tracks cards rather than people it omits the unbanked population (though we suspect that their travel may be relatively unaffected by HSR) and over-counts those who use multiple cards on a given trip. Reassuringly, Appendix A2 demonstrates high correlations between our payments-based mobility data and that obtained from Baidu Migration, a source based on China's leading navigation service.

[5]The extent of station coverage grew substantially from 2013-2016 but, where necessary, we use temporal interpolation to fill in missing values at the city-time level. Our results are similar when using alternative interpolation methods and to dropping observations with missing home pollution measures.

[6]For reference, the U.S. Environmental Protection Agency's daily standard is 35 μg/m^3 (US EPA 2024).

these cities j on date t. Further, we define *travelers' pollution* as

$$TP_{i,t} = \sum_{j \neq i} \frac{N_{ij,t}}{\sum_{j \neq i} N_{ij,t}} P_{j,t}, \tag{1}$$

which is the pollution exposure of the typical traveler from city i on date t. We can then decompose EPE into the roles played by stayers and travelers as follows

$$EPE_{i,t} = \underbrace{N_{ii,t} \times P_{i,t}}_{\text{Total stayers' exposure}} + \underbrace{NT_{i,t} \times TP_{i,t}}_{\text{Total travelers' exposure}}, \tag{2}$$

where $N_{ii,t}$ and $NT_{i,t} = \sum_{j \neq i} N_{ij,t}$ denote the number of stayers and travelers from city i, respectively. That is, EPE is a mix of the pollution exposure that stayers experience at home and the exposure that travelers experience elsewhere.

While the existing literature studies pollution exposure via the first term in (2), the novelty of our analysis is a focus on the second. In particular, we quantify how China's HSR investments changed EPE by altering both the extent of travel (i.e. $NT_{i,t}$) and the extent to which the pollution exposure of travelers (i.e. $TP_{i,t}$) is different from the pollution they leave behind at home (i.e. $P_{i,t}$).

HSR Network.—Figure A1 illustrates the expansion of China's HSR network. While the country had no HSR at the beginning of the 21st century, the network had linked all major cities by 2020, spanning over 38,000 km. We gather information on the opening dates of HSR stations from official reports and hence construct an indicator variable HSR_{it} that equals one if city i has at least one HSR station on date t.

Descriptive Evidence.—As a first exploration of the empirical distinction between travelers' pollution ($TP_{i,t}$) and home pollution ($P_{i,t}$), Figure 1 plots a nonparametric regression of the former on the latter, separately for city-day observations with and without HSR access. Both regression lines cross the 45-degree line from above, a reflection of spatial regression to the mean. But importantly, the HSR line lies strictly below the non-HSR line, and even more so on days when $P_{i,t}$ is extreme.[7] Our regression analysis below explores the difference-in-differences (DID) analog of this difference and its causes.

II. Empirical Framework

We now develop a procedure that aims to estimate the causal effect of HSR access on experienced pollution exposure.

A. *Intensive vs. Extensive Margins of Travel*

Expression (2) quantifies $EPE_{i,t}$ at date t for residents whose home city is i. We now evaluate the change in this measure between a date $t = 0$ (without HSR access) and $t = 1$ (with it), and further imagine that such changes are evaluated relative to a valid control city that did not gain access. Our regression analysis below uses such DID comparisons to isolate causal effects of HSR

[7]Table A2 adds additional quantification of this descriptive evidence.

access. Abstracting from any change in city populations (i.e. $\Delta N_{ii} = -\Delta NT_i$), the (exact) change in experienced exposure can be written as

$$\text{(3)} \qquad \Delta EPE_i = \underbrace{N_{ii,t=0} \times \Delta P_i}_{\text{Direct Effect}} + \underbrace{NT_{i,t=0} \times \Delta TP_i}_{\substack{\text{Intensive Margin} \\ \text{of Travel}}} + \underbrace{(TP_{i,t=1} - P_{i,t=1}) \times (\sum_j N_{ij,t=1}) \times \Delta TS_i}_{\text{Extensive Margin of Travel}},$$

where "Δ" denotes the change in any variable and $TS_{i,t}$ denotes the share of residents from city i who are traveling on date t (i.e. $TS_{i,t} = NT_{i,t}/\sum_j N_{ij,t}$).

The first term in equation (3) captures the possibility that pollution itself has changed in city i. This "direct" effect of HSR—which could occur via an impact of HSR on the composition of the city's economy or the extent of vehicle traffic—is not the focus of our analysis.[8] Instead, we seek to understand the scope for HSR to change EPE via changes in residents' travel patterns, which could mitigate or exacerbate the effects of pollution even without any causal effect of HSR on pollution emissions itself. In what follows we therefore refer to effects of HSR on EPE that deliberately set the direct effect in (3) to zero.[9]

One such change in travel patterns appears in the second term of equation (3). We refer to this as a change in the *intensive margin* because it reflects changes in the pollution exposure of an average traveler (i.e. TP_{it} via equation (1)) who may visit a different set of cities after the HSR connection. By contrast, the final term captures the *extensive margin* by measuring increases in the share of travelers; such changes will be good for reducing EPE_i if travelers' pollution is lower than home pollution in the post-HSR period (i.e. whenever $TP_{i,t=1} < P_{i,t=1}$) and bad otherwise.

In the analysis that follows, we first use DID analysis to estimate the causal effect of HSR access on the two responses in equation (3): ΔTP_i, which drives the intensive margin, and ΔTS_i, which drives the extensive margin. The calculations in Section III.C then evaluate the consequences of these responses for EPE by following the logic of equation (3).

B. Regression Framework

Intensive Margin: Travelers' Exposure.—We begin with an analysis of the intensive margin—that is, how travelers' pollution $TP_{i,t}$ was affected by China's HSR system. To do so we estimate a DID specification of the form:

$$\text{(4)} \qquad TP_{i,t} = \beta HSR_{i,t} + \gamma P_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t},$$

where $HSR_{i,t}$ indicates whether city i is connected to the HSR network at time t, and μ_i and δ_t denote city and time fixed effects, respectively. Our baseline additionally controls for the extent of pollution in city i $P_{i,t}$, though this turns out to matter little (see Section III).

A recent econometrics literature has shown that the traditional two-way fixed effects (TWFE) estimator may be biased in the presence of heterogeneous or dynamic treatment effects in a staggered rollout design such as ours (Goodman-Bacon 2021; de Chaisemartin and d'Haultfoeuille 2020; Sun and Abraham 2021). To address this issue, we report estimates based on the CSDID

[8] Recent research by Lee et al. (2023) and Chang et al. (2021) documents evidence of such phenomena in China.

[9] An extended version of such "direct" effects could also occur via ΔTP_i, in the intensive margin effect, if HSR access in city i causes a change in destination cities' pollution as well. However, we find no evidence for this in specifications that replace actual TP_{it} with an analog that holds travel behavior constant in each city's pre-HSR era.

procedure developed by Callaway and Sant'Anna (2021) throughout our analysis.[10] This procedure is computationally intractable for long panels, so we take the time period t to be a month and return to the question of temporal aggregation below.[11]

Our coefficient of interest, β, captures—when using a method such as CSDID—the average treatment effect of HSR connection on travelers' pollution exposure. Its identification relies on the assumption that the specific timing of a city's connection to the HSR network is unrelated to the pollution in destinations (weighted by the extent of their importance in the travel patterns of the city's residents). In practice, the timing of HSR network expansion is a complex process involving multiple entities and numerous steps—siting, land purchase, financing, physical construction, etc.—which means that the opening date of many HSR stations was dictated by idiosyncratic factors.[12] Consistent with this context, the event-study estimates plotted in panel (a) of Figure 2 display no clear differential trends in TP before HSR connection.

To ensure that our estimate of β can be appropriately combined with the decomposition in equation (3), our intensive margin regressions are weighted by the number of travelers in each city period. Further, to accommodate arbitrary serial correlation in the error term, standard errors are clustered at the city level.

Extensive Margin: Outbound Trips.—Our study of the extensive margin proceeds analogously: we estimate equation (4) but where the dependent variable is $TS_{i,t}$, the share of city i's residents who are traveling at time t.[13] The identification assumption in this case—namely, that the timing of the HSR connection is uncorrelated with changes in the travel share that would have happened in the absence of HSR connection—is arguably less plausible for the extensive margin than for the intensive margin, since the placement of HSR could directly respond to the growing demand for intercity travel, or there could be contemporaneous shocks that affect both placement and travel demand. To assess the potential non-random placement of HSR stations, we perform a battery of analyses. As with the case of the intensive margin analysis discussed above, we probe the plausibility of the parallel trends assumption in an event-study analysis (Panel b of Figure 2) and see no evidence of differential trends. In addition, we find reassuringly similar results when using the synthetic DID method, with staggered treatments, developed by Arkhangelsky et al. (2021).

Finally, our weighting procedure for the extensive margin analyses applies weights based on the number of residents in each city-period. This implies that the β estimate will reflect the effect of HSR on an average resident's propensity to travel.

[10] This approach computes the cohort-time average treatment effects separately for each treatment cohort relative to not-yet-treated control units and aggregates them to obtain the average treatment effect.

[11] We construct monthly variables as weighted averages of daily measures. For example, for month t, $TP_{i,t} \equiv \frac{1}{|\mathscr{D}_t|} \sum_{d \in \mathscr{D}_t} \sum_{j \neq i} \frac{N_{ij,d}}{\sum_{k \neq i} N_{ik,d}} \times P_{j,d}$, where $\mathscr{D}_t$ is the set of days in month t and $N_{ij,d}$ and $P_{j,d}$ are values on day d.

[12] Most HSR lines in our sample began construction in 2005 or right after 2008, directly following the passage of the Mid-to-Long Term Railway Network Plan in 2004 and its revision in 2008. However, their completion dates were spread out over multiple years, depending on factors such as engineering difficulties, bridge/tunnel ratios, political considerations, accidents, and other forces that could expedite or delay projects. For example, part of the Wuhan-Guangzhou line involved building the Wuhan Tianxingzhou Yangtze River Bridge, which has the world's longest combined road and rail span and took nearly six years to build. The construction progress was also halted for over a year after the Wenzhou train crash in July 2011.

[13] As above, for any time period t we construct $TS_{i,t}$ as the weighted average of daily observations within t.

III. Did HSR Reduce Experienced Pollution Exposure?

We now report estimates of the effect of HSR access on experienced pollution exposure, appealing to the decomposition of equation (3) into intensive and extensive margins.

A. *Intensive Margin*

Panel (a) in Table 1 presents results on the intensive margin from estimating equation (4).[14] Our baseline specification, presented in column (1), uses a dependent variable $TP_{i,t}$ that measures the share of days in month t on which travelers from city i experience "extreme" pollution at their destinations, and where the "extreme" cutoff is defined as exceeding 75 μg/m^3.[15] This specification therefore measures the effect of a city joining the HSR network on the likelihood of travelers from that city experiencing extreme pollution. The reported CSDID coefficient ($\widehat{\beta}$) of -0.031 implies that for the typical city—holding constant its own home pollution level—the likelihood that travelers from the city experience extreme pollution is reduced by approximately 14% (the population-weighted average of the dependent variable is 0.211) as a result of HSR connection. This CSDID estimate is an average over all dynamic treatment effects, but the underlying dynamic responses are illustrated in the event-study plot in Figure 2 (panel a).

Columns (2) to (5) of Table 1 present results that explore the sensitivity of this baseline estimate to alternative specifications and samples. Column (2) uses a continuous pollution measure (the log of $PM_{2.5}$ concentration) as the dependent variable and estimates that the HSR connection reduces travelers' pollution exposure by about 3.6%.[16] Column (3) demonstrates that our estimates of the effect of HSR are not materially affected by whether we control for home pollution or not. Finally, columns (4) and (5) investigate the role of outliers—the former specification excludes the top and bottom 5% of days in terms of the number of trips for each city and the latter drops holidays and the forty-day travel rush period surrounding the Lunar New Years. The estimated coefficient $\widehat{\beta}$ changes little (from -0.031 to -0.032) across these sample cuts.

Additional robustness checks appear in Appendix Tables A3 and A4. Specifically, Table A3 confirms that the conclusions in our baseline specification are not sensitive to alternative choices about what constitutes "extreme" pollution. And Table A4 confirms that the aggregation of daily pollution readings into a monthly dependent variable appears to be inconsequential.[17]

B. *Extensive Margin*

Panel (b) of Table 1 turns to the extensive margin analog of equation (4). Beginning with our baseline in column (1), the dependent variable is the fraction of a city's residents who engage in outbound trips. We estimate an HSR-connection effect of 0.007, which implies (given a weighted

[14]Our data covers 303 cities over a span of four years from January 2013 to December 2016, though the CSDID procedure drops 55 cities that begin with HSR access, resulting in a total of 11,712 city-month observations.

[15]That is, the measure of pollution $P_{i,d}$ used to construct $TP_{i,d}$ in equation (1) for any day d is an indicator variable for whether that day's average $PM_{2.5}$ measure exceeds 75 μg/m^3.

[16]Figure A4 displays the corresponding event-study figure for the regression in column (2).

[17]Because CSDID analysis is infeasible at the sub-monthly level the robustness checks in Table A4 compare feasible TWFE estimates based on daily, weekly and monthly versions. These estimates are lower than the CSDID estimate—as expected, given the growth in coefficients in the event-study of Figure 2—but their similarity to one another suggests that aggregation bias is not a concern in our setting.

mean value for the dependent variable of 0.25) that HSR access gives rise to a (statistically significant) 3% increase in out-of-city travel for a typical resident. Panel (b) of Figure 2 displays the accompanying event-study figure.

Moving to robustness checks, column (2) reports that our findings are very similar if we apply the synthetic DID procedure of Arkhangelsky et al. (2021). And the remainder of panel (b) confirms that dropping the home pollution control, dropping extreme travel days, and dropping holiday periods—in columns (3) to (5) respectively—is just as inconsequential for the extensive margin estimated here as it was for the intensive margin of panel (a).

C. Total Effect on EPE and Accompanying Health Benefits

We now use the decomposition presented in Section II.A to combine the estimated intensive and extensive margins. This delivers an estimate of the effect of HSR access on total experienced pollution exposure (holding constant the direct effect, which is not our focus). We also discuss the health benefits that may occur as a result of the implied changes in pollution exposure.

Beginning with the intensive margin of equation (3), this requires an estimate of the response in travelers' exposure from HSR connection (ΔTP_i) and an estimate of the number of travelers on a typical day in the pre-HSR era ($NT_{i,t=0}$). For the former, we use the estimate in column (2) of Table 1, based on the continuous measure of $PM_{2.5}$, which implies that HSR caused a reduction of daily exposure of 2.5 $\mu g/m^3$ of $PM_{2.5}$ for a typical traveler from an HSR-connected city.[18] For the latter, recognizing that our card data may not capture the correct ratio of travel to non-travel transactions (see Section I), we instead use the best available official statistics to estimate that on average in 2013 at least 79.5 million residents of the cities that were HSR-connected by 2016 were engaged in intercity travel daily on average.[19] Putting these two numbers together then implies that, due to the intensive margin, HSR access lead to a permanent reduction of $PM_{2.5}$ exposure by 2.5 $\mu g/m^3$ for at least 79.5 million people.

Turning to the extensive margin, according to equation (3) this contribution can be arrived at by multiplying two numbers: (i) the causal response of the share of the population that travels (i.e. ΔTS_i for the average resident); and (ii) the extent to which travelers' and home pollution differ for the average resident in the post-HSR era (i.e. $\sum_i(\sum_j N_{ij,t=1})(TP_{i,t=1} - P_{i,t=1})$). The first of these numbers—from column (3) of Table 1—is a 3% increase in TS.[20] For the second number, we calculate a population-weighted, post-HSR, average value of $TP_{i,t=1} - P_{i,t=1} = 54.37 - 54.36 = 0.01$ $\mu g/m^3$, meaning that pollution levels were slightly higher for travelers than stayers. This implies that, in contrast to the intensive margin effect, the extensive margin actually contributes to

[18]The log specification coefficient estimate translates to a 3.6% reduction. At the sample mean of daily $PM_{2.5}$ concentration in 2013 (69 $\mu g/m^3$) this is a daily reduction of 2.5 $\mu g/m^3$.

[19]According to the Ministry of Transportation, Chinese travelers made a total of 29 billion one-way intercity trips using commercial transportation (i.e., not including self-driving) in 2008. Our data shows that travelers in HSR-connected cities accounted for 80% of all travelers. Assuming a trip duration of 2.5 days based on our data, we derive a lower-bound estimate of 79.5 million intercity travelers per day among residents of HSR-connected cities (i.e., 29 billion $\times \frac{1}{2} \times 80\% \times \frac{1}{365} \times 2.5$). This implies approximately 6% of China's population travelling outside their home city on an average day, a number that is consistent with travel volumes inferred from mobile phone records. For example, using an updated version of the data in Li et al. (2023), we calculate that in Guangdong province (whose 21 cities span a range of development similar to that in our full sample) 9.4% of resident phone users engaged in inter-city travel on an average day in March 2024.

[20]This use of our baseline specification for the component ΔTS_i in the final term of equation (3) ignores any potential correlation between pollution and travel responses along the extensive margin. As discussed in Section IV.A, such a correlation is ambiguous *a priori* and estimated to be negligible.

an *increase* in EPE. However, the combined effect of numbers (i) and (ii) is very small: an extra 0.95 million travelers experiencing the 0.01 $\mu g/m^3$ of $PM_{2.5}$ exposure per day associated with travel on average.[21] The intensive-margin benefit discussed above is therefore almost four orders of magnitude larger in absolute value than these extensive-margin damages.

We have seen that the total effect of HSR on EPE is essentially driven entirely by the intensive margin effect, so it amounts to 79.5 million people enjoying 2.5 $\mu g/m^3$ less $PM_{2.5}$ exposure on average on any given day. To put this magnitude in context, one can ask what health benefits we might expect to arise from such reduced exposure. This is clearly a speculative exercise, and one that does not provide a complete cost-benefit analysis as it omits other potential benefits (such as reduced morbidity) as well as costs (such as the private and social costs of travel). With these caveats in mind, using estimates from the literature on mortality effects of pollution exposure in China, we estimate that the intercity travel enabled by the HSR network (as of 2016) gives rise to 21.3 million life-years saved.[22]

IV. Underlying Mechanisms

The previous section quantified the effect of China's HSR expansion on experienced pollution exposure. In this section we provide three sets of evidence about how such effects materialized.

A. Avoidance Behavior

Our analysis so far has examined the impact of HSR access on the exposure to pollution of those who travel, as well as the propensity to travel altogether, on average across all types of pollution conditions at home. Motivated by the descriptive patterns in Figure 1, we now estimate the extent to which these phenomena are affected by the degree of home-city pollution. To do so we estimate triple-difference variants of equation (4), alongside its extensive margin equivalent, that include an interaction between HSR and the level of home pollution.[23]

Beginning with the intensive margin, panel (a) of Table 2 reports estimates from this augmented specification. Columns (1) and (2) do so using daily observations, but vary whether the level effect of HSR access—no longer the effect of interest in these interaction specifications—is modeled as homogeneous or heterogeneous across cities. Columns (3) and (4) are analogous but examine monthly aggregates. All four cases are similar: the interaction between $P_{i,t}$ and HSR is negative, statistically significant, and large enough to account for essentially all of the baseline effect discussed in Section III. That is, the days on which HSR access appears to reduce travelers' pollution exposure tend to be those on which home pollution is extreme. One possible explanation for this finding is that HSR access leads to longer travel distances, particularly on days that are especially polluted at home (and hence also nearby). Columns (5) and (6) of Table 2, panel (a), examine such

[21] The above estimate of 79.5 million travelers on a typical day in the base period, combined with an estimated 3% growth due to HSR, implies an extra 0.95 million travelers.

[22] Ebenstein et al. (2017) estimate that one additional $\mu g/m^3$ of sustained exposure to PM_{10} reduces life expectancy by 0.064 years in China and Zhou et al. (2016) estimate that one $\mu g/m^3$ increase in ambient $PM_{2.5}$ corresponds to a 1.67 $\mu g/m^3$ increase in PM_{10}. To the extent that the former impact of PM_{10} reflects a stable mixture of impacts due to both PM_{10} and $PM_{2.5}$ concentrations, and that the latter relationship between PM_{10} and $PM_{2.5}$ is stable in our sample, 79.5 million travelers experiencing a sustained reduction in $PM_{2.5}$ of 2.5 $\mu g/m^3$ daily would lead to 21.3 million life-years saved (i.e., 79.5 million $\times$ 2.5 $\times$ 1.67 $\times$ 0.064).

[23] These results draw on TWFE because the level and interaction terms are not jointly estimable with CSDID.

behavior change directly by estimating how the (log) average outbound trip distance is affected by HSR access. We see that average distances do indeed increase after HSR access, and only on days when the home city suffers from extreme pollution.[24] Taken together, the results in panel (a) of Table 2 are therefore suggestive of one form of avoidance behavior in which travelers adjust their destinations to reduce pollution exposure, and HSR facilitates this activity.

Columns (1) to (4) in Panel (b) of Table 2 conduct an analogous investigation into the extensive margin response. In this case the interaction term between the HSR dummy and the home extreme dummy is close to zero and never statistically significant. This result is perhaps unsurprising given that home pollution may both encourage travel to cleaner cities as reported in Chen et al. (2021) and discourage it (as part of a broader behavior of remaining indoors) as documented in Barwick et al. (2024).

B. Composition of Destination Cities vs. Travel Distance

As we have seen, HSR access enabled further travel, but it also expanded the set of travel destinations for any given distance. Which of these mechanisms contributes most to the reductions in experienced pollution exposure presented earlier? To answer this question we construct a counterfactual version of $TP_{i,t}$ but in which the actual pollution at each destination $P_{j,t}$ in equation (1) is replaced with that of a randomly chosen city within a similar ($\pm$25 km) distance band from origin i. We then re-estimate specification (4) but using this counterfactual version of $TP_{i,t}$ as the dependent variable (and repeat this 100 times to reduce simulation noise). This exercise therefore purges any potential for pollution avoidance via the selection of favorable destinations that are equally far away. What remains after removing such a selection mechanism is therefore purely the result of traveling further.

The results from this exercise (in Column 1 of Table 3) show that HSR access had a considerably smaller effect on this counterfactual version of $TP_{i,t}$ (-0.011) than it did on actual $TP_{i,t}$ (-0.031). Put differently, we know from Table 2 that HSR allowed travel to farther destinations especially on days when home pollution was extreme. But the fact that distant cities tend to have lower pollution than the home city on such days—a natural consequence of a decaying spatial autocorrelation in pollution—accounts for only about one-third of our estimated effect.

C. Adaptation Horizons

A separate question of interest concerns the time horizon at which HSR-driven experienced pollution exposure reductions occur. At one end of the spectrum, residents may use HSR-enabled travel to evade short-run surprises in ambient pollution; at the other end, HSR access may facilitate a long-run form of travel adjustment that simply avoids destinations during predictably undesirable time windows. To quantify such possibilities, we proceed analogously to the study of travel distance in Section IV.B, but where our counterfactual measure of $TP_{i,t}$ is now computed using counterfactual travel patterns rather than counterfactual pollution levels at the destination. Specifically, for each pair of locations and day, we calculate travel shares on the basis of long-run averages (separately for each home city's pre- and post-HSR regimes) so as to purge $TP_{i,t}$ of any form of shorter-run adaptive behavior. Column 2 of Table 3 reports the results of this procedure. The HSR coefficient estimate falls to -0.024, which means that the long-term adjustment accounts for 77% of HSR's contribution towards reducing EPE.

[24] Additional visualizations of these effects are presented in Figure A3.

D. Summary

Putting together the findings in this section, we conclude that the bulk of the reduction in experienced pollution exposure caused by the HSR network can be explained by the way that better travel connections enabled travelers to access an expanded set of roughly equidistant destinations that are predictably less polluted.

V. Conclusion

This study has examined how improved transportation infrastructure in China has altered the air pollution exposure experienced by Chinese residents via the promotion of spatial mobility. Leveraging a unique data merge—between geospatial data on the daily locations of travelers, hourly pollution readings throughout the country, and the massive expansion of high-speed railway connections—we find that the HSR network had by 2016 caused changes in travel that reduced experienced pollution exposure, holding constant any direct effects on pollution itself, by an amount that would translate into 21.3 million life-years saved when applying recent estimates about the mortality effects of pollution in China. Perhaps surprisingly, this effect is not primarily driven by the extensive margin (more individuals leaving their home cities), by short-run adaptation (individuals avoiding unexpectedly large pollution realizations), or by the expanded distances involved in typical trips. Instead, it results predominantly from individuals changing their travel destinations to cities that offer predictably cleaner air and are no further from home.

These findings contribute to both the understanding of the role of novel forms of adaptation to environmental shocks and the ability of transportation infrastructure investments to enhance such adaptation. One important goal for future research is to embed the behavioral responses we have tracked into a choice-theoretic model so as to compare the potential health benefits of travel to the costs of such behavior, as well as to measure the extent to which the benefits we have identified are enjoyed only by specific populations (such as the relatively affluent).

REFERENCES

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager, "Synthetic difference-in-differences," *American Economic Review*, 2021, *111* (12), 4088–4118.

Arlt, Wolfgang Georg, "As Smog Hits China, Chinese Tourists Seek Fresh Air On Pollution Free Holidays," *Forbes*, Jan 2017.

Athey, Susan, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt, "Estimating experienced racial segregation in US cities using large-scale GPS data," *Proceedings of the National Academy of Sciences*, 2021, *118* (46), 1–9.

Balboni, Clare, "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities," 2021. Working Paper.

Banzhaf, H. Spencer and Randall P. Walsh, "Do People Vote with Their Feet? An Empirical Test of Tiebout," *American Economic Review*, 2008, *98*, 843–863.

Barreca, Alan, Karen Clay, Olivier Deschênes, Michael Greenstone, and Joseph S Shapiro, "Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the twentieth century," *Journal of Political Economy*, 2016, *124* (1), 105–159.

Barwick, Panle Jia, Shanjun Li, Liguo Lin, and Eric Zou, "From Fog to Smog: the Value of Pollution Information," *American Economic Review*, 2024, *115* (5), 1–45.

Baum-Snow, Nathaniel, "Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960-2000," *American Economic Review*, 2010, *100* (2), 378–82.

Bayer, Patrick, Nate Keohane, and Christopher Timmins, "Migration and Hedonic Valuation: The Case of Air Quality," *Journal of Environmental Economics and Management*, 2009, *58*, 1–14.

Bernard, Andrew B, Andreas Moxnes, and Yukiko U Saito, "Production networks, geography, and firm performance," *Journal of Political Economy*, 2019, *127* (2), 639–688.

Black, Richard, Stephen RG Bennett, Sandy M Thomas, and John R Beddington, "Climate change: Migration as adaptation," *Nature*, 2011, *478* (7370), 447.

Burke, Marshall B, Edward Miguel, Shanker Satyanath, John A Dykema, and David B Lobell, "Climate robustly linked to African civil war," *Proceedings of the National Academy of Sciences*, 2010, *107* (51), E185–E185.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Carleton, Tamma A and Solomon M Hsiang, "Social and economic impacts of climate," *Science*, 2016, *353* (6304), aad9837.

Carleton, Tamma, Amir Jina, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, Andrew Hultgren, Robert E Kopp, Kelly E McCusker, Ishan Nath et al., "Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits," *The Quarterly Journal of Economics*, 2022, *137* (4), 2037–2105.

Chang, Zheng, Chenghao Deng, Fenjie Long, and Longfei Zheng, "High-speed rail, firm agglomeration, and PM2. 5: Evidence from China," *Transportation Research Part D: Transport and Environment*, 2021, *96*, 102886.

Chen, Shuai, Paulina Oliva, and Peng Zhang, "The effect of air pollution on migration: Evidence from China," *Journal of Development Economics*, 2022, *156*, 102833.

_ , Yuyu Chen, Ziteng Lei, and Jie-Sheng Tan-Soo, "Chasing Clean Air: Pollution-Induced Travels in China," *Journal of the Association of Environmental and Resource Economists*, 2021, pp. 59–89.

Chen, Yuyu, Avraham Ebenstein, Michael Greenstone, and Hongbin Li, "Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy," *Proceedings of the National Academy of Sciences*, 2013, *110*, 12936–12941.

Cruz, José-Luis and Esteban Rossi-Hansberg, "The Economic Geography of Global Warming," *Review of Economic Studies*, 2024, *91* (2), 899–939.

Currie, Janet and Matthew Neidell, "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience," *Quarterly Journal of Economics*, 2005, *120*, 1003–1030.

de Chaisemartin, Clément and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken, "Temperature shocks and economic growth: Evidence from the last half century," *American Economic Journal: Macroeconomics*, 2012, *4* (3), 66–95.

Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif, "The mortality and medical costs of air pollution: Evidence from changes in wind direction," *American Economic Review*, 2019, *109* (12), 4178–4219.

Deschênes, Olivier and Enrico Moretti, "Extreme weather events, mortality, and migration," *The Review of Economics and Statistics*, 2009, *91* (4), 659–681.

_ , Michael Greenstone, and Jonathan Guryan, "Climate Change and Birth Weight," *American Economic Review*, May 2009, *99* (2), 211–217.

Donaldson, Dave, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng Zhou, "New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy," *Proceedings of the National Academy of Sciences*, 2017, *114*, 10384–10389.

Faber, Benjamin, "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Freeman, Richard, Wenquan Liang, Ran Song, and Christopher Timmins, "Willingness to pay for clean air in China," *Journal of Environmental Economics and Management*, 2019, *94*, 188–216.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Steffano Polloni, and Matthew A. Turner, "Subways and Urban Air Pollution," *American Economic Journal: Applied Economics*, 2022, *14* (1), 164–96.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Hsiang, Solomon M, Marshall Burke, and Edward Miguel, "Quantifying the influence of climate on human conflict," *Science*, 2013, *341* (6151), 1235367.

Ito, Koichiro and Shuang Zhang, "Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China," *Journal of Political Economy*, 2020, *128*, 1627–1672.

Kahn, Matthew, *Climatopolis: How Our Cities will Thrive in Our Hotter Future*, Basic Books, 2010.

Khanna, Gaurav, Wenquan Liang, A. Mushfiq Mobarak, and Ran Song, "The Productivity Consequences of Pollution-Induced Migration in China," *American Economic Journal: Applied Economics*, 2024. Forthcoming.

Landrigan, Philip, Richard Fuller, Nereus J R Acosta, Olusoji Adeyi, Maureen Cropper, Alan Krupnick, Michael Greenstone, and et al., "The Lancet Commission on pollution and health," *The Lancet*, 2018, *391* (10119), 462–512.

Lee, Wang-Sheng, Trang My Tran, and Lamont Bo Yu, "Green infrastructure and air pollution: Evidence from highways connecting two megacities in China," *Journal of Environmental Economics and Management*, 2023, *122*, 102884.

Li, Teng, Panle Jia Barwick, Yongheng Deng, Xinfei Huang, and Shanjun Li, "The COVID-19 pandemic and unemployment: Evidence from mobile phone data from China," *Journal of Urban Economics*, 2023, *135*, 103543.

Lin, Yatang, "Travel costs and urban specialization patterns: Evidence from China's high speed railway system," *Journal of Urban Economics*, 2017, *98*, 98–123.

Moretti, Enrico and Matthew Neidell, "Pollution, health, and avoidance behavior evidence from the ports of Los Angeles," *Journal of Human Resources*, 2011, *46* (1), 154–175.

Sharkov, Damien, "Smog in China Prompts Tide of Tourism Fleeing 'Airpocalypse:' Report," *Newsweek*, Dec 2016.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

US EPA, "Reconsideration of the National Ambient Air Quality Standards for Particulate Matter," 2024. https://www.govinfo.gov/content/pkg/FR-2024-03-06/pdf/2024-02637.pdf.

Zhang, Junjie and Quan Mu, "Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks," *Journal of Environmental Economics and Management*, 2018, *92*, 517–536.

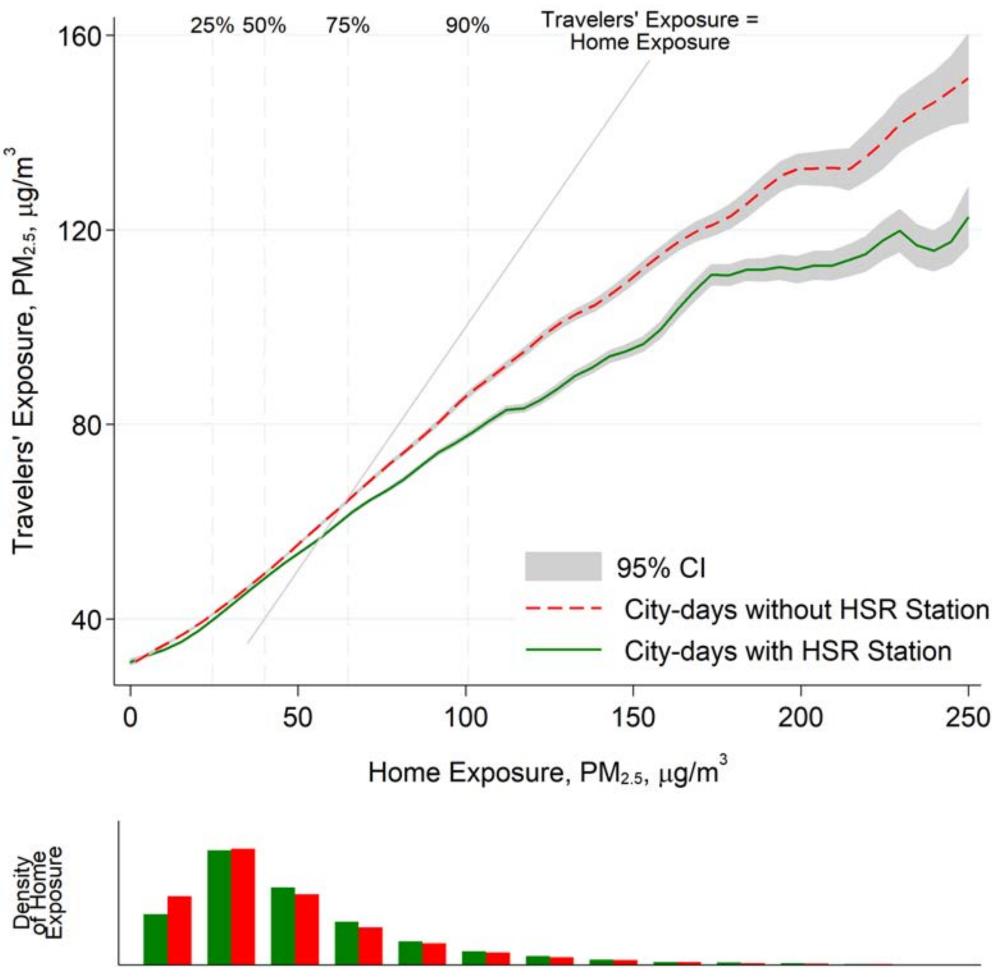
Zheng, Siqi and Matthew E Kahn, "China's bullet trains facilitate market integration and mitigate the cost of megacity growth," *Proceedings of the National Academy of Sciences*, 2013, *110* (14), E1248–E1253.

_ **and** _ , "A new era of pollution progress in urban China?," *Journal of Economic Perspectives*, 2017, *31* (1), 71–92.

Zhou, Xuehua, Zhaoyu Cao, Yujie Ma, Linpeng Wang, Ruidong Wu, and Wenxing Wang, "Concentrations, correlations and chemical species of PM2. 5/PM10 based on published data in China: potential implications for the revised particulate standard," *Chemosphere*, 2016, *144*, 518–526.

Zivin, Joshua Graff and Matthew Neidell, "Temperature and the allocation of time: Implications for climate change," *Journal of Labor Economics*, 2014, *32* (1), 1–26.

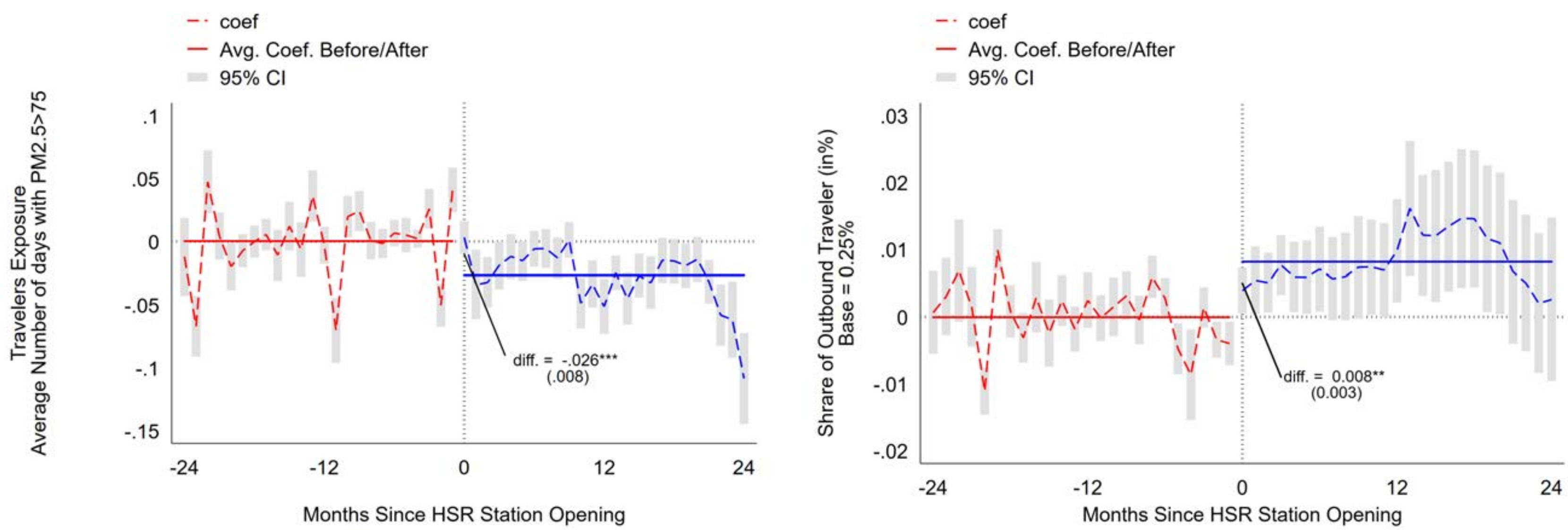
FIGURE 1. TRAVELERS' VS. HOME POLLUTION EXPOSURE



Notes: The top figure plots travelers' average exposure to $PM_{2.5}$ (*y*-axis) against the $PM_{2.5}$ level in the home city (*x*-axis), separately for city-days with HSR access (green line) and city-days without HSR access (red dash line). Both lines are local polynomial regressions weighted by an Epanechnikov kernel with optimal bandwidth. The gray area is the 95% confidence interval. Gray vertical lines mark the 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of the $PM_{2.5}$ distribution. The bottom figure displays the distribution of daily $PM_{2.5}$ at home cities, separately for city-days with HSR access (green bars) and city-days without HSR access (red bars).

FIGURE 2. HSR IMPACTS ON EXPERIENCED POLLUTION EXPOSURE

(a) INTENSIVE MARGIN: TRAVELERS' EXPOSURE (b) EXTENSIVE MARGIN: OUTBOUND TRAVEL



Notes: This figure plots event study coefficient estimates (red and blue dashed lines) relative to the month of HSR station openings for travelers' pollution exposure (left figure, intensive margin) and the travelers' population share (right figure, extensive margin) with monthly observations, using the CSDID method of Callaway and Sant'Anna (2021). The estimated labeled "diff." is calculated with reference to the shown event window only, so it can differ from those computed in Table 1.

TABLE 1—THE EFFECT OF HSR ON EXPERIENCED POLLUTION EXPOSURE

	Panel (a): Intensive margin: travelers' exposure to pollution extremes				
	(1)	(2)	(3)	(4)	(5)
				Subsamples	
	Baseline	Continuous	No control for home pollution	Excl. 5% days with most/least trips	Excl. holidays and 40-day travel rush
	$PM_{2.5} > 75$	$\log(PM_{2.5})$	$PM_{2.5} > 75$	$PM_{2.5} > 75$	$PM_{2.5} > 75$
HSR_{it}	-0.031	-0.036	-0.034	-0.032	-0.032
	(0.008)	(0.011)	(0.008)	(0.008)	(0.009)
N	11,632	11,632	11,712	11,632	11,143

	Panel (b): Extensive margin: share of residents who travel				
	(1)	(2)	(3)	(4)	(5)
				Subsamples	
	Baseline	Synthetic DID	No control for home pollution	Excl. 5% days with most/least trips	Excl. holidays and 40-day travel rush
HSR_{it}	0.007	0.008	0.007	0.007	0.008
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
N	11,632	11,664	11,712	11,632	11,146

Notes: This table examines the effect of HSR expansion on travelers' exposure (the intensive margin) in Panel (a) and on the share of outbound trips (the extensive margin) in Panel (b) via Equation (4). The estimation is conducted using the CSDID method (Callaway and Sant'Anna 2021). The unit of observation is a city-month. Month-of-sample FEs and city FEs are included in all regressions. The coefficient estimates are calculated as a weighted average of city-specific ATTs, where the weights are the average number of travelers from each city for Panel (a) and residents in Panel (b). The dependent variable in Columns (1), and (3) to (5) in Panel (a) indicates the likelihood of travelers from a city experiencing $PM_{2.5}$ pollution over 75 $\mu g/m^3$. The mean of the dependent variable, weighted by the number of outbound travelers for each city-day, is 0.211. The dependent variable in Column (2) is the logarithm of travelers' pollution exposure. HSR_{it} is an indicator variable for access to an HSR station. We control for home pollution as the fraction of days with extreme pollution ($PM_{2.5} > 75\mu g/m^3$) at home city i for each month (in Columns (1), (4) and (5)) or the logarithm of monthly average of $PM_{2.5}$ (in Column (2)). Columns (4) and (5) examine the robustness of the intensive and extensive margins of the HSR effect as reported in Column (1) using different sub-samples. The dependent variable in Panel (b) is the share of travelers (proxied by the ratio of cards with out-of-town transactions over the total number of active cards). It is winsorized at the 1% level, and the mean is 0.25%. Columns (2) report the unweighted ATT across post-treatment periods from synthetic DID estimations based on Arkhangelsky et al. (2021). The specifications for Columns (3) to (5) are the same as the corresponding specifications in Panel (a). Standard errors are clustered at the city level.

TABLE 2—HETEROGENEITY OF HSR EFFECTS WITH RESPECT TO HOME EXPOSURE

	Panel (a): intensive margin					
	Travelers' exposure to pollution extremes				log(avg. trip dist.)	
	(1)	(2)	(3)	(4)	(5)	(6)
HSR_{it}	0.009	Absorbed	0.006	Absorbed	0.001	Absorbed
	(0.008)	-	(0.006)	-	(0.027)	-
Home Extreme	0.189	0.189	0.239	0.240	-0.025	-0.024
	(0.012)	(0.012)	(0.021)	(0.021)	(0.027)	(0.027)
Home Extreme $\times$ HSR_{it}	-0.081	-0.081	-0.065	-0.065	0.038	0.036
	(0.023)	(0.023)	(0.014)	(0.015)	(0.020)	(0.019)
Sample	Daily	Daily	Monthly	Monthly	Monthly	Monthly
N	330,801	330,801	11,054	11,049	10,982	10,977
R^2	0.81	0.81	0.94	0.94	0.91	0.91

	Panel (b): extensive margin			
	Share of outbound trips			
	(1)	(2)	(3)	(4)
HSR_{it}	0.0127	Absorbed	0.0129	Absorbed
	(0.0080)	-	(0.0077)	-
Home Extreme	-0.0004	0.0002	-0.0044	-0.0029
	(0.0015)	(0.0013)	(0.0067)	(0.0064)
Home Extreme $\times$ HSR_{it}	0.0009	0.0001	0.0058	0.0028
	(0.0018)	(0.0016)	(0.0059)	(0.0055)
Sample	Daily	Daily	Monthly	Monthly
N	330,819	330,819	11,054	11,049
R^2	0.78	0.79	0.83	0.84

Notes: This table examines the heterogeneity of HSR impacts with respect to home exposure, the triple-difference variant of Table 1. The main variable of interest is the interaction of the HSR dummy with *Home Extreme*, which denotes whether home city *i* experiences extreme pollution ($\text{PM}_{2.5} > 75\mu g/m^3$) on a given day for daily specifications in Columns (1) and (2), and denotes the fraction of days with extreme pollution at home city *i* in a given month for monthly specifications in other columns. See Table 1 for the definition of the dependent variables in Columns (1)-(4) and other regressors. The dependent variable in Columns (5) and (6) in Panel (b) is the average distance for outbound trips (in logarithm) originating from city *i* for each month. All regressions use the TWFE method and are weighted by home cities' number of outbound travelers in period *t* in Panel (a) and residents in period *t* in Panel (b). Columns (1) and (2) include day-of-month FEs while other columns include month-of-sample FEs. City FEs are included in all regressions. Columns (2) and (4) allow for the city-specific main effect of HSR (β_{2i}) and hence the HSR dummy is absorbed. Standard errors are clustered at the city level.

TABLE 3—THE EFFECT OF HSR ON TRAVELERS' POLLUTION EXPOSURE: UNDERLYING MECHANISMS

	Counterfactual Travelers' Exposure Measure	
	Only distance factor (1)	Only longer horizon adaptation (2)
HSR_{it}	-0.011 (0.005)	-0.024 (0.007)
N	11,620	11,632

Notes: This table explores the intensive margin of the HSR effect as reported in Column (1) of Table 1 along different underlying mechanisms. In Column (1), the dependent variable is a counterfactual travelers' exposure measure that is constructed by replacing destination cities' environmental conditions with the concurrent conditions in cities of a similar distance to the origin, $\widehat{TP}_{it} = \sum_{j \neq i} \widehat{P}_{jt} \cdot \frac{N_{ijt}}{\sum_{k \neq i} N_{ikt}}$, where $\widehat{P}_{jt}$ is the pollution level of a randomly-chosen city that is similarly distant from the original city as city j in day t. We repeat the analysis 50 times and the number of observations varies between 11,620 and 11,632 due to the randomization process with the minimum number of observation reported in the table. Column (1) reports the mean and the standard deviation of the coefficient estimates from the 50 repetitions. In Column (2), the dependent variable is a counterfactual travelers' exposure measure that is constructed using the pre-HSR and post-HSR daily average travel flows from an origin city to a destination on days: $\widetilde{TP}_{it} = \sum_{j \neq i} P_{jt} \cdot \frac{\overline{N}_{ij}^{r(t)}}{\sum_{k \neq i} \overline{N}_{ik}^{r(t)}}$, where $\overline{N}_{ij}^{r(t)}$ denotes the average number of travelers from city i visiting destination j during that city's pre-HSR or post-HSR connection regime r, and $r(t)$ indicates which regime date t belongs to. The results in both columns use the CSDID method (Callaway and Sant'Anna 2021). The unit of observation is city-month. All regressions control for home pollution. Standard errors are clustered at the city level.

Appendices

Additional Data Description and Evidence

Cross-Validating UnionPay Trip Data with Baidu Data.— To examine the quality of our intercity travel flows based on card transaction data, we compare our data with intercity mobility data from Baidu Migration by Baidu Maps, China's leading provider of digital map and online navigation services. Baidu migration data report population inflows from the top 100 origin cities and outflows to the top 100 destination cities between January 21 and March 23 in 2019 and between January 10 and March 15 in 2020.[25] The data are based on the location of over 600 million users of Baidu's location-based services, hence providing good coverage and accuracy.

Panels (a) and (b) in Figure A2 show that the two measures of intercity travels have high correlations, with a R^2 of 0.79 for both outflows and inflows. Panel (c) depicts the coefficient estimates of a gravity equation where the travel frequencies between city pairs (in logarithm) from the two data sources are separately regressed on flexible distance bins between the origin and destination cities. The coefficient estimates based on the UnionPay and Baidu data are very close to each other across all distance bins. These validation exercises confirm the quality of our travel flow measures based on card transaction data.

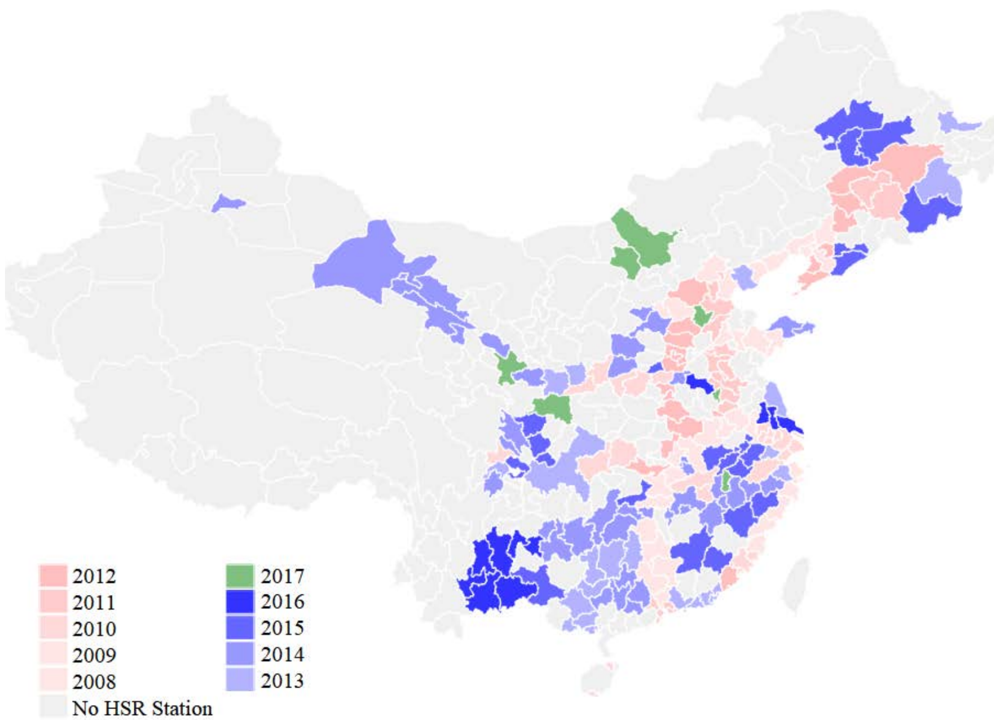
Summary Statistics.— Table A1 presents summary statistics for key variables used in this analysis. There are a total of 330,819 city-day observations for $PM_{2.5}$ readings from 2013 to 2016. Some cities did not have $PM_{2.5}$ monitoring stations in the beginning of our sample, which explains $PM_{2.5}$'s lower number of observations. The average daily $PM_{2.5}$ is 52.19 $\mu g/m^3$ across all city-days (unweighted) and 53.79 $\mu g/m^3$ when weighted by the number of travelers originating from each city-day. This is considerably higher than U.S.'s daily standard of 35 $\mu g/m^3$. About 19% of the city-days also exceeded China's national ambient air quality standard for daily concentration of $PM_{2.5}$ at 75 $\mu g/m^3$ (which became effective in 2016).

$PM_{2.5}$ varies considerably in our sample period, with a standard deviation of 44.51 $\mu g/m^3$ and an inter-quartile range of 40.3 $\mu g/m^3$. About 80% of the variation comes from day-to-day changes within a city, while the remaining 20% arises from differences across cities (i.e., within vs. between variation in a panel setting).

[25] Source: http://qianxi.baidu.com/. Baidu Migration data aim to help understand population flows during the Chinese New Year so the data duration centers around the time of Chinese New Year. The data are only available since 2019.

APPENDIX FIGURES AND TABLES

FIGURE A1. EXPANSION OF THE HSR NETWORK

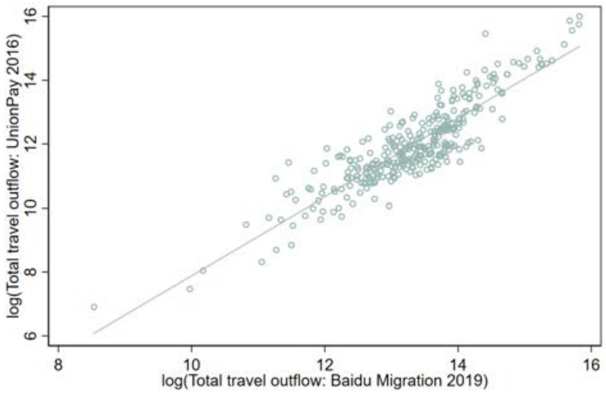


Year	In 2010	'11	'12	'13	'14	'15	'16	'17
HSR network								
# Cities Added	55	16	21	23	40	25	11	7
# Cities in Network	55	71	92	115	155	180	191	198

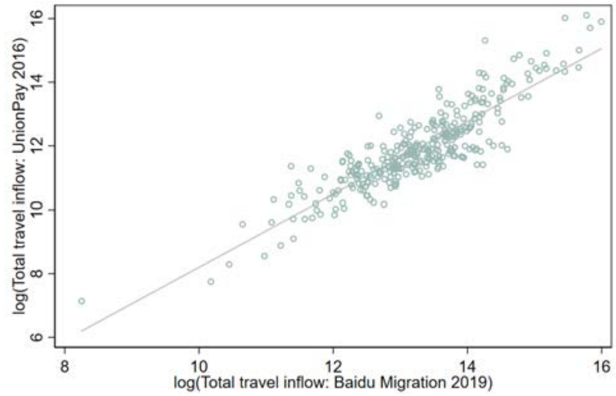
Notes: This map depicts the rollout of the HSR network from 2010 to 2017. Different colors represent the year when a city is first connected to the HSR network. Cities in gray did not have HSR connections by the end of 2017. The number of cities in the network is calculated at the end of each year.

Figure A2. Validation using Baidu Migration Data

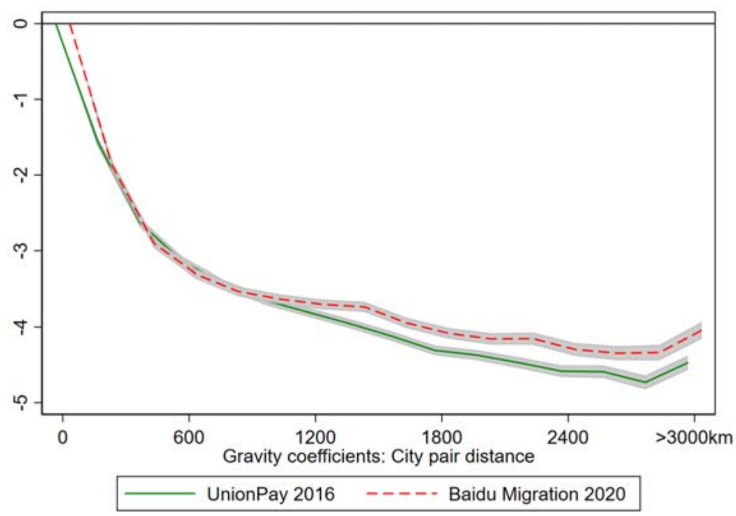
(a) Population outflow



(b) Population inflow



(c) Gravity equation estimates



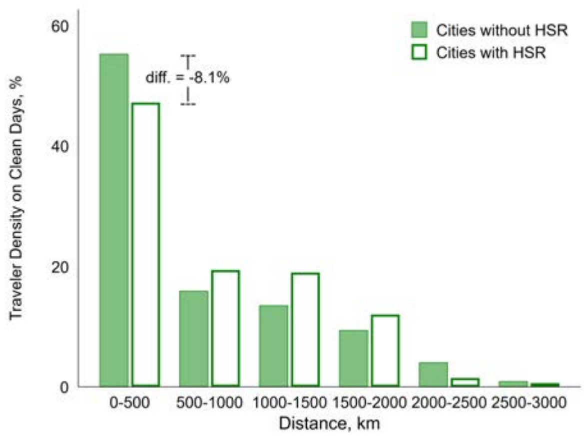
Notes: Panel (a) plots the (log) total number of outbound trips originating from each city using UnionPay data on the y-axis against that using the Baidu Migration index on the x-axis. Panel (b) plots the (log) total number of arriving trips from each city using UnionPay data on the y-axis, against that using the Baidu Migration index on the x-axis. Each dot represents a city. Panel (c) plots the estimates of β_k from the gravity equation below, based on city-pair bilateral trips from UnionPay data and the Baidu Migration Index, respectively:

$$ln(N_{ij}) = \sum \beta_k I_k + \alpha_i + \gamma_j + \varepsilon_{ij},$$

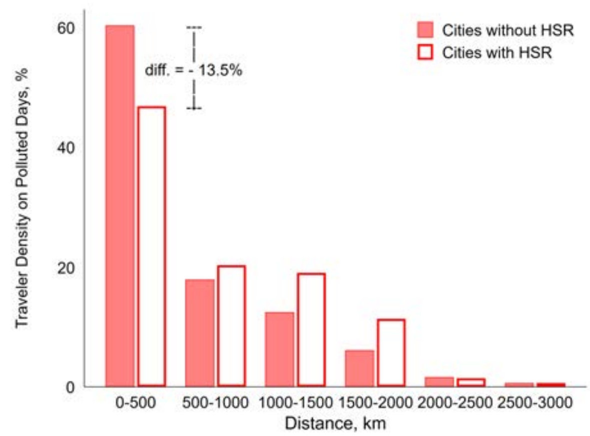
where N_{ij} is the aggregate trips between i and j in 2016 for the Unionpay data and in 2019 for Baidu data. I_k stands for fifteen 200 km interval distance bands k (0-200 km is chosen as the base group) and α_i and γ_j are origin and destination city fixed effects.

FIGURE A3. TRIP DISTANCE FOR CITY-DAYS DURING CLEAN AND POLLUTED DAYS

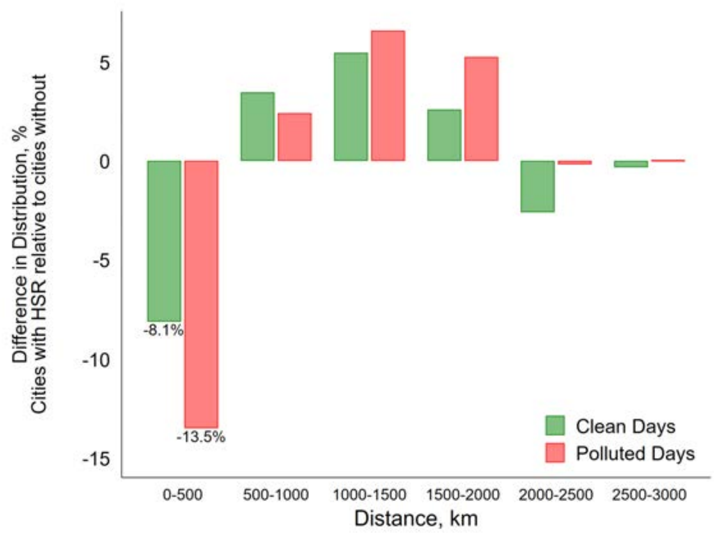
(a) TRIP DISTANCES DURING CLEAN DAYS



(b) TRIP DISTANCES DURING POLLUTED DAYS

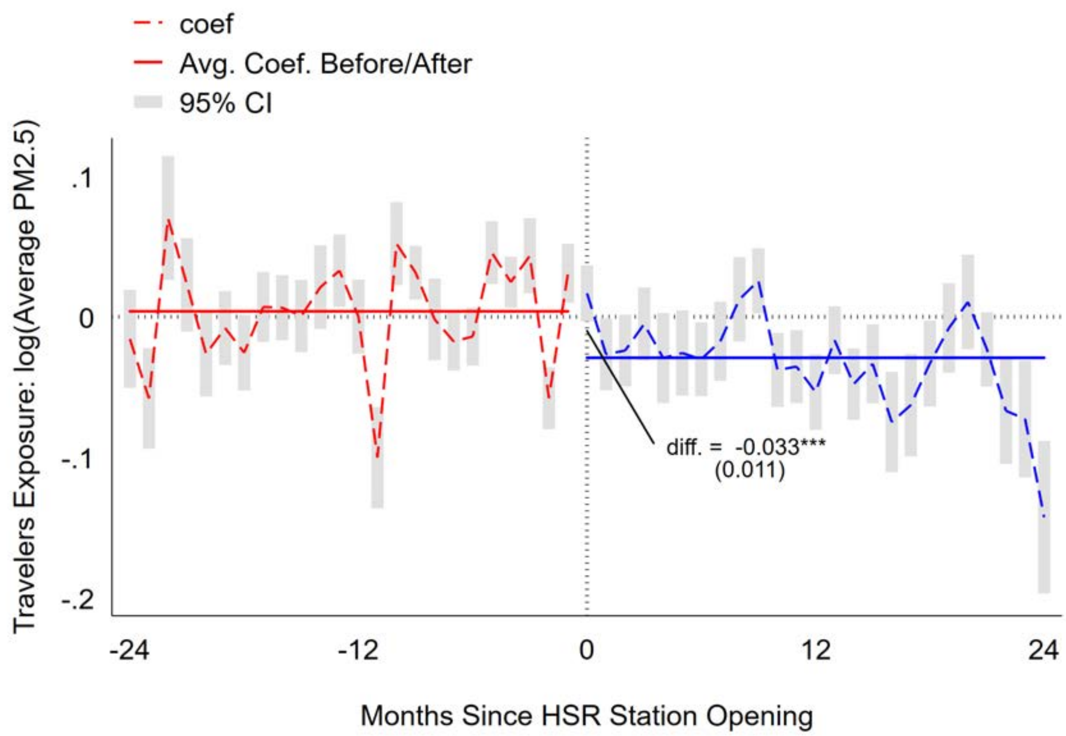


(c) DIFFERENCE IN TRIP DISTANCES



Notes: Panels (a)-(b): histograms for trip distances for cities without and with HSR access during clean and polluted ($PM_{2.5} > 75\mu g/m^3$) days. Panel (c) plots differences in trip length density between cities with HSR access and cities without, on clean days (green bars) and polluted days (red bars). Travelers from cities with HSR access are more likely to visit distant destinations and especially so when home cities experience extreme pollution.

FIGURE A4. ROBUSTNESS: HSR ACCESS ON TRAVELERS' EXPOSURE USING $\log(PM_{2.5})$



Notes: The figure plots coefficient estimates (red and blue dashed lines) by month relative to HSR station openings for travelers' pollution exposure measured by log(PM2.5). The average treatment effects over the full post-treatment period, weighted by the number of travelers, is -3.6%, as reported in Column (2) of Table 1. The event study analysis uses the algorithm of Callaway and Sant'Anna (2021).

TABLE A1—SUMMARY STATISTICS

	Mean	Std. Dev.	Min	Max	Number of Obs.
$PM_{2.5}$, $\mu g/m^3$ (Unweighted)	52.19	44.51	0	1782.98	330,819
$PM_{2.5}$, $\mu g/m^3$ (Weighted)	53.79	44.64	0	1782.98	330,819
$\mathbb{1}\{PM_{2.5} > 75\}$ (Unweighted)	0.20	0.40	0	1	330,819
$\mathbb{1}\{PM_{2.5} > 75\}$ (Weighted)	0.21	0.40	0	1	330,819
Number of travelers	189.65	500.12	1	15,013	486,209
Share of Outbound Cards	0.0025	0.0012	0.00008	0.182	486,184
HSR	0.42	0.49	0	1	486,209

Notes: This table reports summary statistics of the city-daily panel dataset. The variable $\mathbb{1}\{PM_{2.5} > 75\}$ takes the value 1 if a city's daily average $PM_{2.5}$ concentration is greater than $75\mu g/m^3$ and 0 otherwise. Weights refer to the number of travelers originating from each city-day. The variable *HSR* takes the value 1 if a city has an HSR station in operation on a given day and 0 otherwise.

TABLE A2—TRAVELERS' TRAVEL PATTERNS AND POLLUTION EXPOSURE

(a) FOR TRAVELERS FROM CITY-DAYS WITH HSR

		Flow shares by pollution quintile at destination					Avg. Exposure of PM$_{2.5}$ ($\mu g/m^3$) at	
		Clean 1	2	3	4	Dirty 5	Home	Dest.
Home in	Clean 1	42.0%	23.9%	15.2%	10.9%	8.0%	16.4	35.1
	2	24.3%	29.3%	21.4%	14.3%	10.7%	28.4	41.9
	3	15.5%	21.1%	27.5%	21.8%	14.1%	41.3	49.1
	4	11.3%	14.1%	21.8%	30.3%	22.5%	60.1	58.7
	Dirty 5	7.7%	9.8%	14.0%	21.7%	46.9%	121.4	84.4

(b) FOR TRAVELERS FROM CITY-DAYS WITHOUT HSR

		Flow shares by pollution quintile at destination					Avg. Exposure of PM$_{2.5}$ ($\mu g/m^3$) at	
		Clean 1	2	3	4	Dirty 5	Home	Dest.
Home in	Clean 1	38.5%	25.8%	16.6%	11.3%	7.8%	16.2	35.6
	2	22.2%	27.7%	23.0%	16.4%	10.7%	28.4	42.9
	3	13.9%	20.7%	27.0%	23.4%	14.9%	41.2	50.2
	4	9.1%	12.5%	21.1%	31.9%	25.4%	60.1	61.8
	Dirty 5	5.2%	7.4%	10.9%	21.2%	55.2%	116.9	92.9

(c) DIFFERENCE BETWEEN PANEL (a) AND PANEL (b)

		Difference in flow shares by pollution quintile					Diff. in PM$_{2.5}$ ($\mu g/m^3$) at	
		Clean 1	2	3	4	Dirty 5	Home	Dest.
Home in	Clean 1	3.5%	-1.8%	-1.4%	-0.4%	0.2%	0.19	-0.49
	2	2.1%	1.6%	-1.5%	-2.1%	0.0%	-0.03	-1.01
	3	1.6%	0.4%	0.4%	-1.6%	-0.9%	0.12	-1.08
	4	2.2%	1.5%	0.8%	-1.7%	-2.8%	0.00	-3.06
	Dirty 5	2.5%	2.4%	3.0%	0.4%	-8.4%	4.56	-8.59

Notes: This table complements Figure 1 and illustrates that travelers from HSR cities are more likely to visit cleaner cities than those from non-HSR cities. Each row represents the shares of travel to destination cities with different pollution quintiles, conditioning on the home city-day's pollution quintile. Panel (a) refers to travelers from city-days with HSR, Panel (b) refers to travelers from city-days without HSR, and Panel (c) presents the difference between the two. The quintile cutoffs for daily PM$_{2.5}$ are 23, 34, 49, and 75 $\mu g/m^3$. The weighted average level of PM$_{2.5}$ for cities with HSR access (i.e., home exposure) is 54.36 $\mu g/m^3$. The weighted average level of PM$_{2.5}$ exposure for travelers from cities with HSR access (i.e., traveler exposure) is 54.37 $\mu g/m^3$.

TABLE A3—INTENSIVE MARGIN WITH ALTERNATIVE MEASURES OF POLLUTION EXPOSURE

	(1)	(2)	(3)	(4)	(5)
	Traveler's likelihood of experiencing $PM_{2.5}$ >Cutoff				
Cutoff ($\mu g/m^3$)	50	55	60	70	80
HSR_{it}	-0.036	-0.034	-0.034	-0.032	-0.025
	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
N	11,632	11,632	11,632	11,632	11,632
Dep. Var. Average	0.40	0.35	0.31	0.24	0.19

Notes: This table examines the robustness of the intensive margin of the HSR effect shown in column (1) of Table 1 to alternative definitions of pollution extremes. The dependent variable is the likelihood that travelers experience extreme pollution ($PM_{2.5}$ >Cutoff) as defined in equation 1, with the cutoffs taking values 50, 55, 60, 70, and 80 $\mu g/m^3$ in Columns (1) to (5), respectively. HSR_{it} is an indicator variable for access to an HSR station. The results in all columns use the CSDID method (Callaway and Sant'Anna 2021). The unit of observation is a city-month. Month-of-sample FEs and city FEs are included in all regressions. We also control for home pollution as the fraction of days with extreme pollution extreme ($PM_{2.5} > 75\mu g/m^3$) in the home city i for each month in all regressions. Standard errors are clustered at the city level.

TABLE A4—THE EFFECT OF HSR AT THE INTENSIVE MARGIN: TWFE REGRESSIONS

	Intensive margin – travelers' likelihood of experiencing pollution extremes ($PM_{2.5} > 75\mu g/m^3$)		
	(1)	(2)	(3)
HSR_{it}	-0.009	-0.009	-0.010
	(0.005)	(0.005)	(0.005)
Home Extreme	0.120	0.165	0.186
	(0.027)	(0.025)	(0.023)
Sample	Daily	Weekly	Monthly
N	330,801	47,406	11,054
R^2	0.80	0.90	0.94

Notes: This table reports the intensive margin of the HSR effect from two-way fixed effects (TWFE) regressions via equation (4). The specification is the same as Column (1) in Table 1. The unit of observation is a city-day in column (1), a city-week in column (2), and a city-month in column (3). HSR_{it} is an indicator variable for access to an HSR station. *Home Extreme* is an indicator for extreme pollution ($PM_{2.5} > 75\mu g/m^3$) at home city i on day t for daily samples and the fraction of days with extreme pollution for weekly or monthly regressions. We drop cities with missing home pollution information and retain always-connected cities in our sample, which explains the difference in the number of observations with Column (1) in Table 1. The regressions are weighted by home cities' number of outbound travelers in period t. All regressions include time(day, week, or month)-of-sample FEs and city FEs. Standard errors are clustered at the city level.