Can Transportation Networks Facilitate Adaptation to Environmental Extremes?

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

The social costs of environmental extremes hinge critically on society's ability to adapt. Based on transaction records from China's card payment network, we compile daily travel flows and estimate how the country's high-speed railway (HSR) system facilitates the use of intercity travel as a means of adaptation to extreme air pollution. We find that HSR expansion reduced realized pollution exposure—that is, exposure based on travelers' actual locations when pollution occurs—by an amount that implies a mortality benefit of 8.5 million life-years saved. This occurs primarily due to travelers' changing towards destinations with predictably cleaner air.

JEL: O18, Q53, Q54, R41

Keywords: Adaptation, Pollution, Transportation Infrastructure

Environmental degradation and climate change present escalating risks.[1] The social costs of extreme environmental conditions hinge crucially on human adaptability. While long-term migration can be an effective strategy for responding to a spatially changing environment (Banzhaf and Walsh 2008; Deschênes and Moretti 2009; Kahn 2010; Freeman et al. 2019; Desmet et al. 2021), such mobility also comes with substantial costs. In contrast, shorter-term travel may offer a more feasible means of adaptation to the adverse effects of local extremes. For example, in China, the emergence of "haze-avoidance tourism" and "smog refugees" (Arlt 2017; Sharkov 2016; Chen et al. 2021) exemplifies this form of adaptation to the country's lethal air pollution levels. Can modern transportation systems facilitate such adaptive behavior?

This study provides what is, to our knowledge, the first analysis of how improved trans-

* 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, Joseph Shapiro, and Matthew Turner for comments, to Nahim Bin Zahur and Ziye Zhang for excellent research assistance, and to Minwei Tang for data support.

[1]Existing work extensively documents the various impacts of pollution and climate change on economic growth (Nordhaus 2006; Dell et al. 2012; Carleton and Hsiang 2016), social stability (Burke et al. 2010; Hsiang et al. 2013), and health (Deschênes et al. 2009; Currie and Neidell 2005; Chen et al. 2013; Landrigan et al. 2018), among many other aspects.

portation infrastructure can change exposure to adverse environmental conditions. To do so, we create a unique dataset of daily travel flows between all city-pairs in China, derived from credit and debit card transactions within the world's largest inter-bank payment network (UnionPay). Combining this dataset with high-frequency pollution readings throughout the country, we build novel measures of *realized pollution exposure* (RPE)—that is, the actual exposure of Chinese residents on the basis of their daily location choices. We then evaluate how the 2013-16 expansion of the country'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—enabled a change in travel behavior that reduces pollution exposure.

Our analysis starts with a conceptual framework that decomposes the impact of HSR on residents' RPE, holding the national distribution of pollution emissions constant, into effects that work on the extensive and intensive margins of travel responses. The extensive margin pertains to individuals traveling more often to other cities, potentially affecting RPE if typical destinations offer different air quality. The intensive margin examines the role of improved travel options in altering the air quality of the destinations that even a fixed number of travelers may choose. Both effects are of unknown magnitude, and their signs are not even clear given that highly polluted locations may nevertheless be attractive travel destinations for a host of other reasons. We therefore quantify the contributions of both margins by estimating the impacts of city-level HSR access on the share of outbound trips and the average pollution exposure of travelers, employing a difference-in-differences (DID) approach supported by event-study analysis that is consistent with the requisite assumption that the timing of HSR access was exogenous.

We uncover two main findings. First, the effect of HSR expansion on the intensive margin of behavior contributes to a substantial reduction in RPE: 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. To put this in context, a reduction of this size translates (when using recent estimates of the mortality effects of pollution in China) into 8.5 million life-years saved. Second, we find that even though HSR causes a sizable expansion in the number of travelers, the effects of this extensive margin response actually increase RPE because, on average, travelers' destinations are slightly dirtier than their home cities. However, this effect is many 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 RPE.

The final component of our analysis asks what types of mechanisms 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 ones that are predictably clean at times when travelers' own cities are predictably dirty (rather than depending on shorter-term deviations in conditions).

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 (Graff Zivin and Neidell 2014; Moretti and Neidell 2011; Barwick et al. 2020), increased defensive investments (Barreca et al. 2016; Ito and Zhang

2020; Zheng and Kahn 2017; Zhang and Mu 2018), and altered migration (Banzhaf and Walsh 2008; Bayer et al. 2009; Deschênes and Moretti 2009; Kahn 2010; Black et al. 2011; Chen et al. 2017; Freeman et al. 2019; Balboni 2021; Cruz and Rossi-Hansberg 2021). Our study is unique in focusing on shorter-term travel rather than longer-term migration, as well as the way that transport infrastructure investments affect this form of adaptation.

Second, we contribute to the literature on quantifying the effects of transportation networks on, for example, enhanced mobility of people and goods (Donaldson 2018; Bernard et al. 2019), deeper market integration (Baum-Snow 2010; Zheng and Kahn 2013; Faber 2014; Lin 2017), and easier knowledge sharing (Agrawal et al. 2017; Dong et al. 2020). Drawing upon the rapid expansion of China's HSR system and employing high-frequency data on intercity travel and pollution, our paper adds to this literature by illustrating a new form of benefit from transportation investments.

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, each described here.[2]

Intercity Travel Flow.—We construct daily bilateral passenger flows across cities from the universe of credit and debit card transactions conducted through the UnionPay network from January 2013 to December 2016.[3][4] We have access to a 1% random sample of cards, which records all of the transactions made by 27 million cards during 2013-2016. We focus on offline transactions, where the cardholder is physically present at the merchant's location, to track the movement of card users.[5] We define an inter-city trip as having occurred whenever the city of a card's offline transaction differs from that of its "home" city.[6] As a result, we arrive at an estimate of the number of residents from city i who travel to location j on date t, denoted $N_{ij,t}$.[7]

HSR Network.—Figure A1 illustrates the expansion of China's HSR network. While the country had no HSR at the beginning of 21st century, the total length of HSR reached

[2]Summary statistics are presented in Table A1, with further discussion in Appendix A.

[3]As the sole interbank payment network in China, UnionPay tracked 34 trillion *yuan* ($4.9 trillion) of purchases annually at this time, which amounts to over 40% of retail consumption. Electronic 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]Cities correspond to prefectures, which include both a city's urban core and its rural periphery.

[5]Online spending accounts for less than 5% by the number of transactions and 3% by the value.

[6]A card's home city at date t is its most common location within a rolling 12-month window centered on date t. We examine robustness in Section III.B.

[7]A limitation of this source is that it tracks cards rather than people. It therefore omits the unbanked population, though we suspect that their travel may be relatively unaffected by HSR access. It also over-counts those with multiple cards, but would only over-count travel when multiple cards are used on a given trip. With these concerns in mind, in Appendix A, we cross-validate the quality of our travel flow data with alternative intercity mobility data from Baidu Migration, China's leading provider of digital map and navigation services. Flows in these two sources exhibit a high correlation and similar spatial patterns.

38,000 km by 2020, linking all of its major cities. 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.

Air Pollution Data.—Since 2013, China's Ministry of Ecology and Environment (MEE) has used ground-level monitoring stations at over 1,000 sites and reported hourly air quality data on its website (Barwick et al. 2020).[8] 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}. For example, 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, 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.[9]

Realized 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. Given any measure of pollution $P_{i,t}$ we define the *realized pollution exposure* of city i on date t as

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

which simply weights the pollution in each city $P_{j,t}$ by the number of residents of city i who are in these cities j on date t. Further, we define *travelers' pollution* as

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

which is the pollution exposure of the typical traveler from city i on date t.[10]

We then split up *RPE* into the roles played by stayers and travelers as follows

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

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. Our analysis below focuses on how transportation infrastructure investments can change the second of these terms—the contribution to pollution exposure played by the extent of travel (i.e. $NT_{i,t}$) and the extent to which the pollution exposure of the typical

[8] The number of such stations increased from 1003 to 1615 in 2015. Where necessary, we use temporal interpolation to fill in missing values at the city-time level, and discuss robustness to alternatives in Section III.B.

[9] For reference, the U.S. Environmental Protection Agency's maximum allowed value is 35 μg/m^3.

[10] Our baseline construction of $TP_{i,t}$ calculates traveler-weighted averages among non-missing $P_{j,t}$ values, but we revisit this in Section III.B.

traveler (i.e. $TP_{i,t}$) is different from home pollution (i.e. $P_{i,t}$).

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 $TP_{i,t}$ on $P_{i,t}$, separately for city-day observations with and without HSR access. Importantly, the HSR regression line lies strictly below the non-HSR line, and even more so on days when $P_{i,t}$ is extreme.[11] Table 1 quantifies this phenomenon. Here we tabulate the matrix of conditional $P_{i,t}$-versus-$TP_{i,t}$ quintiles, separately for HSR access (panel a) and non-access (panel b), and for the difference in these two matrices (panel c). The last column of panel (c) reports a small HSR-access difference in $TP_{i,t}$ on days when $P_{i,t}$ was relatively low (0.49 μg/m^3) but a large difference (8.59 μg/m^3) when it was high. Our regression analysis below explores the 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 travelers' pollution exposure.

A. Intensive vs. Extensive Margins

Recall that expression (2) quantified $RPE_{i,t}$ at date t for residents whose home city is i. We now evaluate the change in this measure between a date $t = 1$ (with HSR access) and $t = 0$ (without 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 access. Abstracting from any change in city populations (i.e. $\Delta N_{ii} = -\Delta NT_i$), the (exact) change in exposure can be written as

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

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.[12] Instead, we seek to understand the scope for HSR to change RPE via changes in

[11] Of course, both regression lines cross the 45-degree line from above, indicating that on average a particularly polluted day in a resident's home city would be less severe for a typical traveler from that city (and vice versa). This reflects spatial regression to the mean.

[12] Recent research by Lee et al. (2023) and Chang et al. (2021) documents evidence of such phenomena. An extended version of such "direct" effects could also occur via the component ΔTP_i 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 a placebo analog that holds travel behavior constant in each city's pre-HSR era.

residents' travel patterns, which could mitigate or exacerbate the effects of pollution even if actual pollution emissions were held constant.

One 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 changes in the share of travelers; such travel will be good for reducing city i's exposure if travelers' pollution is lower than home pollution in the post-HSR period (i.e. whenever $TP_{i,t=1} > P_{i,t=1}$).

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 *RPE* 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

$$TP_{i,t} = \beta HSR_{i,t} + \gamma P_{i,t} + \mu_i + \delta_t + \varepsilon_{i,t}, \tag{4}$$

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.B).

A recent econometrics literature has argued 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 procedure developed by Callaway and Sant'Anna (2021) throughout our analysis.[13] This procedure is computationally intractable for long panels, so we take the time period t to be a month; section III.B returns to the issue of temporal aggregation.[14]

Our coefficient of interest, β, captures the impacts 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 nu-

[13] 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.

[14] We construct monthly variables as weighted averages of daily measures. For example, for month t, $TP_{i,t} = \frac{1}{T_t}\sum_{d=1}^{T_t}\sum_{j\neq i} \frac{N_{ij,d}}{\sum_{k\neq i} N_{ik,d}} \times P_{j,d}$, where T_t is the number of days in month t and $N_{ij,d}$ and $P_{j,d}$ are values on day d.

merous steps—siting, land purchase, financing, physical construction, etc.—which means that the opening date of many HSR stations was dictated by idiosyncratic factors.[15] Consistent with this context, the event-study estimates plotted in panel (a) of Figure 2 display no clear trend in travelers' pollution exposure before the 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, 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.[16] 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 control and synthetic DID methods, with staggered treatments, developed by Arkhangelsky et al. (2021) and Ben-Michael et al. (2022).

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.

III. Did HSR Reduce Travelers' Pollution Exposure?

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

[15]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.

[16]As above, for any time period t we construct $TS_{i,t}$ as the weighted average of daily observations within t (e.g. monthly $TS_{i,t}$ is the share of residents traveling outside the city, on average, for any given day within the month).

A. *Baseline Results*

Table 2 presents the result of estimating equation (4) and its extensive margin analog.[17] Column (1) begins with the intensive margin response, and where pollution (both in the dependent variable, $TP_{i,t}$, and the control, $P_{i,t}$) is measured as an indicator for "extreme" pollution ($PM_{2.5}$ exceeding 75 μg/m^3). The reported coefficient 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). We estimate $\widehat{\beta} = -0.031$, which is statistically significant at standard levels. This value implies that for the typical city—holding constant its own "home" pollution level, which is always controlled for throughout our analysis—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 result of HSR connection. Column (2) repeats this estimate but using a continuous pollution measure ($PM_{2.5}$ concentration in logarithm); our results are similar when implemented this way.[18]

Column (3) of Table 2 turns to the extensive margin response—that is, the dependent variable in our CSDID regressions is now the fraction of a city's residents engaging in outbound trips. Panel (b) of Figure 2 displays the accompanying event-study figure. The estimated coefficient is 0.007, which implies (given a weighted mean value for this 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.

The estimates in Table 2 are key inputs into our calculations below about how much China's HSR system led to a change in the total amount of realized pollution exposure. We return to those calculations in Section III.C, after first discussing a range of sensitivity checks on the regression analysis performed so far.

B. *Sensitivity Analysis*

We begin with a series of alternative specifications of the intensive margin analysis:

- Table A2 revisits the effect of HSR access on travelers' exposure that is reported in Panel (a) of Table 2, but with the use of alternative cutoff values for what constitutes "extreme" pollution. Columns (1) to (5) use cutoffs that range from 50 to 80 $\mu g/m^3$ for daily average concentration of $PM_{2.5}$. The estimated coefficients are consistently negative and statistically significant. And the estimated effect size implies a percentage reduction in the dependent variable ranging from 10-14%, similar to that in our baseline.

- Table A3 examines the issue of temporal aggregation in our dependent variable. While CSDID analysis at the sub-monthly level is infeasible, we see that daily, weekly and monthly versions of TWFE estimates are reassuringly stable in response

[17] 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.

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

to temporal aggregation.[19] This lends credence to our use of monthly aggregates in the baseline.

- Table A4 investigates three potential concerns about missing pollution observations. While our baseline analysis interpolated $P_{i,t}$ between missing values, column (1) instead drops observations in which $P_{i,t}$ is missing. Column (2) shows that part of the reason for this robustness is that even removing the control for $P_{i,t}$ altogether has little effect on our estimate. Finally, column (3) considers an alternative approach to the missing pollution values inside $TP_{i,t}$.[20] All three of these results are similar to our baseline.

Next, we turn to a set of alternative approaches to the parallel-trends assumption underlying our estimates of the extensive margin response. We repeat this analysis using synthetic control and synthetic DID in Table A5. The estimated effects of HSR on $TS_{i,t}$ are all positive, statistically significant, and similar in magnitude to our baseline estimate.

Finally, Table A6 examines the robustness of both intensive and extensive margin responses to different sample cuts. Because the number of trips is measured by out-of-town card transactions, we drop city-days with extreme values to mitigate potential measurement error. Columns (1) and (2) exclude the top and bottom 5% and 10% of days with the most or least trips for each city. Column (3) drops holidays and the forty-day travel rush period surrounding the Chinese Spring Festival, as travel patterns for these periods might be outliers. Our baseline findings remain similar across these samples.

C. Total Effect on RPE and Accompanying Health Benefits

Having confirmed the robustness of the intensive and extensive margin estimates in Table 2, the final step in our analysis is to use the decomposition presented in Section II.A to combine these two responses. This delivers an estimate of the effect of HSR access on total realized 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, 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 2, based on the continuous measure of $PM_{2.5}$, which amounts to a reduction of daily exposure of 2.5 $\mu g/m^3$ for a typical traveler an HSR-connected city.[21] For the latter, recognizing that our card data may not capture the correct ratio of travel to non-travel (see Section I), we instead use the best available official statistics to estimate that on

[19]The reported estimates are considerably lower than the CSDID estimate, which is unsurprising given the growth in coefficients in the event-study of Figure 2.

[20]To fill in missing values in $P_{j,t}$ (when constructing $TP_{i,t}$), here we apply the same interpolation procedure to $P_{j,t}$ as used for $P_{i,t}$ in our baseline.

[21]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 $\mu g/m^3$.

average in 2013 at least 31.8 million residents of HSR-connected cities were engaged in intercity travel daily.[22] Putting these two numbers together then implies that, due to the intensive margin, the HSR expansion leads to a permanent reduction of $PM_{2.5}$ exposure by 2.5 $\mu g/m^3$ for at least 31.8 million people.

Turning to the extensive margin, this contribution can be arrived at (via equation 3) 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 2—is a 3% increase in TS.[23] For the second number, we calculate a population-weighted, post-HSR, average value of $TP_{i,t=1} - P_{i,t=1} = 54.37 - 54.36 = 0.01$ $\mu g/m^3$. This implies that, in contrast to the intensive margin effect, the extensive margin actually contributes to *greater RPE*. However, the combined effect of numbers (i) and (ii) is relatively 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.[24] 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 *RPE* is essentially driven entirely by the intensive margin effect, so it amounts to 31.8 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 but we believe it may be of interest nonetheless.[25] Using estimates from the literature on mortality effects of pollution exposure in China, we estimate that the intercity travel enabled by the 2013-16 HSR expansion gives rise to 8.5 million life-years saved.[26]

IV. Avoidance Behavior and Underlying Channels

The previous section focused on quantifying the effect of China's HSR system on realized pollution exposure. To understand *how* the effects documented above materialized, we

[22] According to the Ministry of Transportation, Chinese travelers made a total of 29 billion one-way intercity trips in 2008. Our data shows that travelers in HSR-connected cities accounted for 80% of all travelers. Assuming a conservative trip duration of one day, we derive a lower-bound estimate of 31.8 million intercity travelers per day among residents of HSR-connected cities (i.e., 29 billion one-way trips $\times \frac{1}{2} \times$ 80% $\times \frac{1}{365}$).

[23] 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 below, such a correlation is ambiguous *a priori* and indeed estimated to be very small.

[24] The above estimate of 31.8 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.

[25] This calculation is not meant to provide a complete cost-benefit analysis of documented adaptation as it ignores the private and social costs of travel and other potential benefits (such as reduced morbidity).

[26] 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. 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, 31.8 million travelers experiencing a sustained reduction in $PM_{2.5}$ of 2.5 $\mu g/m^3$ daily would lead to 8.5 million life-years saved (i.e., 31.8 million $\times$ 2.5 $\times$ 1.67 $\times$ 0.064).

provide three sets of evidence about underlying mechanisms.

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 and Table 1, we now estimate triple-difference variants of equation (4), alongside its extensive margin equivalent, that include an interaction between HSR and the level of home pollution.[27]

Beginning with the intensive margin, panel (a) of Table A7 reports estimates from this augmented specification. Columns (1) and (2) do so using daily observations, but vary whether the level effect of HSR access is homogeneous or heterogeneous across cities. Column (3) instead examines monthly aggregates. All three 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 $TP_{i,t}$ tend to be those on which $P_{i,t}$ is extreme. One way in which this could arise is illustrated in Figure A3, which documents how HSR access implies longer travel distances, particularly on days that are especially polluted at home (and hence also nearby). This pair of results is suggestive of a form of avoidance behavior in which travelers adjust their destinations to reduce pollution exposure, and that HSR facilitates this activity.[28] Our analysis below digs deeper into the nature of this adaptive behavior.

Panel (b) of Table A7 conducts an analogous investigation into the extensive margin response. In this case the interaction term is close to zero and never statistically significant. This result is perhaps unsurprising given that home pollution may both encourage travel (to cleaner cities) and discourage it (as part of a wider attempt to remain indoors).

B. *Composition of Destination Cities vs. Travel Distance*

HSR access provides the option to both travel further and to travel to a wider set of destinations within any distance radius. Which of these mechanisms contributes most to the intensive-margin responses presented earlier? To answer this question we construct an augmented version of $TP_{i,t}$ but in which the actual pollution at each destination $P_{j,t}$ is replaced with that of a randomly chosen city within a similar (± 25 km) distance band from origin i. We then re-estimate specification (4) but using this augmented 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 panel a of Table 3) show that HSR access had a considerably smaller effect on this augmented version of $TP_{i,t}$ (-0.011) than it did on actual

[27] Here we use TWFE because the level and interaction terms are not jointly estimable with CSDID.

[28] However, the results in Table A7 say only that the spatial correlation between $P_{i,t}$ and $TP_{i,t}$ became weaker after HSR access, which could also have other root causes.

$TP_{i,t}$ (-0.031). Put differently, we know from Figure A3 that HSR allowed travel to more distant destinations, and more so 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 this form of adaptation behavior occurs. At one end of the spectrum, residents may use HSR-driven travel to evade short-run surprises in ambient pollution; at the other end, HSR access may enable a long-run form of travel adjustment that simply avoids predictably undesirable destinations. To quantify this, we proceed analogously to the study of travel distance in Section IV.B, but where our augmented 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-of-year, we calculate travel shares on the basis of long-run averages (done separately for each home city's pre- and post-HSR windows) so as to purge $TP_{i,t}$ of any form of shorter-run adaptive behavior. Panel (b) of Table 3 reports the results of this procedure. The HSR coefficient estimate falls to -0.024, which means that long-term adjustment accounts for 77% of HSR's contribution towards reducing RPE.

Putting together the three findings in this section, we therefore conclude that the bulk of our results can be explained by HSR enabling travelers to access an expanded set of roughly equidistant destinations that are on average less polluted at times when travelers' own cities are heavily polluted.

V. Conclusion

This study has examined how improved transportation infrastructure in China alters realized exposure to the country's lethal air pollution 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 2013-16 HSR rollout reduced realized pollution exposure by an amount that would translate into 8.5 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 city) or by short-run adaptation (individuals avoiding surprisingly large pollution realizations). Instead, it results predominantly from individuals changing their travel destinations to ones offer predictably cleaner air.

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 the extent to which the benefits we have identified are enjoyed only by the relatively affluent.

REFERENCES

Agrawal, Ajay, Alberto Galasso, and Alexander Oettl, "Roads and innovation," *Review of Economics and Statistics*, 2017, *99* (3), 417–434.

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.

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," 2020.

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.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein, "Synthetic controls with staggered adoption," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2022, *84* (2), 351–381.

Bernard, Andrew B, Andreas Moxnes, and Yukiko U Saito, "Production networks, geography, and firm performance," *Journal of Political Economy*, 2019, *127* (2), 000–000.

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.

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

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," 2017. NBER Working Paper No. 24036.

_ , 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," 2021. Working Paper.

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.

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.

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.

Desmet, Klaus, Robert E. Kopp, Scott A. Kulp, Dávid Krisztián Nagy, Michael Oppenheimer, Esteban Rossi-Hansberg, and Benjamin H. Strauss, "Evaluating the Economic Cost of Coastal Flooding," *American Economic Journal: Macroeconomics*, April 2021, *13* (2), 444–86.

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

Dong, Xiaofang, Siqi Zheng, and Matthew E Kahn, "The Role of Transportation Speed in Facilitating High Skilled Teamwork," *Journal of Urban Economics*, 2020.

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.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021.

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.

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.

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.

Nordhaus, William D, "Geography and macroeconomics: New data and new findings," *Proceedings of the National Academy of Sciences*, 2006, *103* (10), 3510–3517.

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.

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.

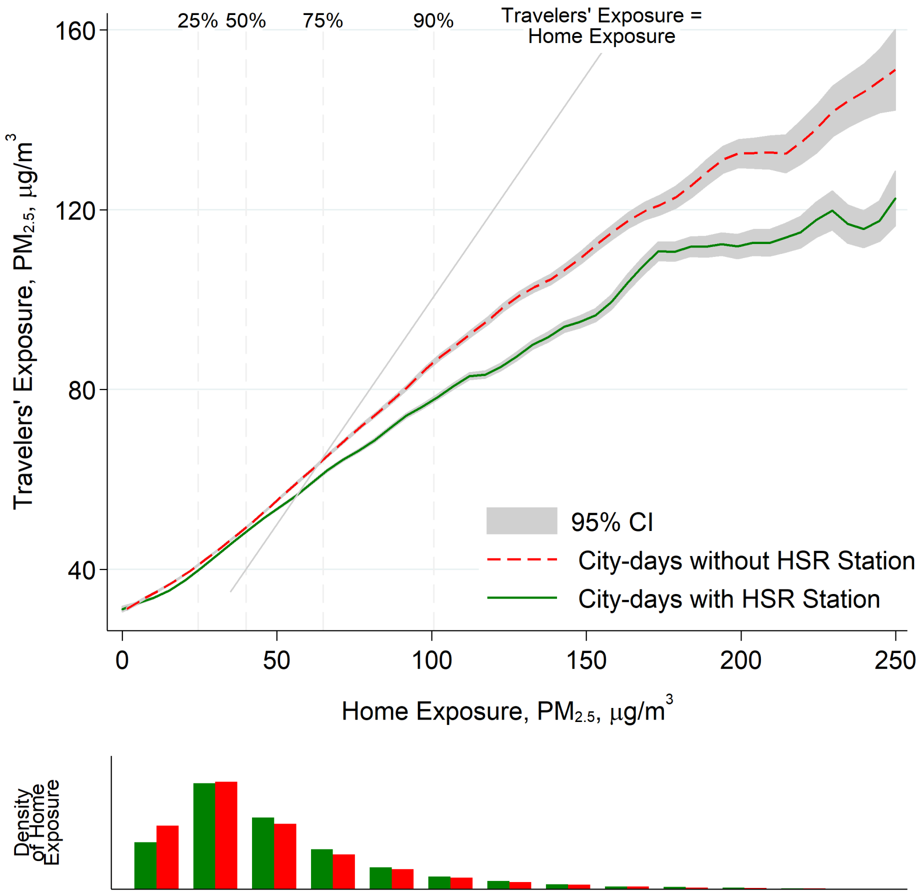
Zheng, Siqi and Matthew E Kahn, "China's bullet trains facilitate market integration and mitigate the cost of megacity growth," *Proceedings of the National Academy of Sciences*, 2013, *110* (14), E1248–E1253.

_ **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' EXPOSURE TO $PM_{2.5}$ VS. HOME EXPOSURE TO $PM_{2.5}$

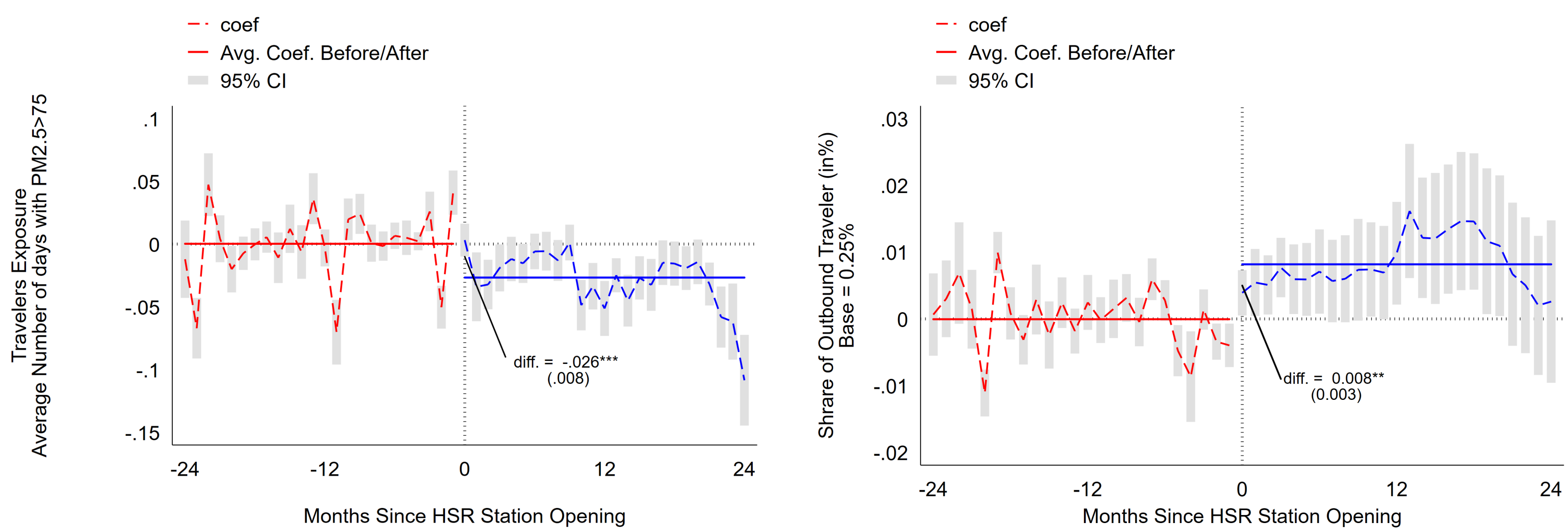


Notes: The top figure plots travelers' average exposure to $PM_{2.5}$ (y-axis) against $PM_{2.5}$ level at the home city (x-axis), separately for city-days with HSR access (green line) and city-days without HSR access (red dash line). Both lines are local polynomial regressions weighted by Epanechnikov 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. Event Study of HSR Impacts on Travelers' Exposure and Outbound Travel

(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 2.

TABLE 1—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 2— THE EFFECT OF HSR AT THE INTENSIVE AND EXTENSIVE MARGIN

	Intensive margin: travelers' exposure to pollution extremes		Extensive margin: share of outbound trips
	(1)	(2)	(3)
	$\text{PM}_{2.5} > 75\mu g/m^3$	$\log(\text{PM}_{2.5})$	
HSR_{it} (β)	-0.031	-0.036	0.007
	(0.008)	(0.011)	(0.004)
Month-of-sample FE	Yes	Yes	Yes
City FE	Yes	Yes	Yes
N	11,632	11,632	11,632

Notes: Columns (1)-(2) examine the HSR effect on travelers' exposure (the intensive margin) via equation (4). The dependent variable in Column (1) is a dummy variable indicating travelers' pollution exposure as defined in Equation (1) to be 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, indicating that the likelihood of travelers being subject to extreme pollution is 21.1%. The dependent variable in Column (2) is the logarithm of traveler's pollution exposure as defined in equation (1). Column (3) explores the HSR effect on the share of outbound trips (the extensive margin). The dependent variable is the share of travelers (proxied by the ratio of cards with out-of-town transactions over the total number of active cards). We winsorize the dependent variables at the 1% level. The dependent variable in column (3) is in percentage points (%) with a mean of 0.25(%). HSR_{it} is an indicator variable for access to HSR stations. The results for all three columns are from CSDID (Callaway and Sant'Anna 2021). The unit of observation is city-month. The reported coefficient is calculated as a weighted average of city-specific ATTs, where the weights are the average number of travelers from each city for Columns (1)-(2) and residents in Columns (3). We control for home pollution in all regressions as the fraction of days with extreme pollution extreme ($\text{PM}_{2.5} > 75\mu g/m^3$) at home city i for each month (in columns 1 and 3) or the logarithm of monthly average of $\text{PM}_{2.5}$ (in column 2). Standard errors are clustered at the city level.

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

	Hypothetical Travelers' Exposure Measure	
Counterfactual Scenarios	Only distance factor	Only longer horizon adaptation
HSR_{it} (β)	-0.011	-0.024
	(0.005)	(0.007)
Month-of-sample FE	Yes	Yes
City FE	Yes	Yes
N	11,620	11,632

Notes: This table explores the intensive margin of the HSR effect as reported in Column (1) of Table 2 along different underlying channels. In Column (1), the dependent variable is a hypothetical 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. The number of observations varies between 11,620 and 11,632 due to the randomization process; the minimum number of observation is reported in the table. Standard errors are estimated from block bootstraps at the city level with 50 repetitions. In Column (2), the dependent variable is a hypothetical 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 with the same day-of-year: $\widetilde{TP_{it}}^{d} = \sum_{j \neq i} P_{jt} \cdot \frac{\overline{N}^{d}_{ijt}}{\sum_{k \neq i} \overline{N}^{d}_{ikt}}$, where $d = 0$ ($d = 1$) denotes city-days without (with) HSR access, $\bar{N}^{d}_{ijt}$ is the average number of travelers from city i visiting destination j pre ($d = 0$) or post ($d = 1$) HSR connections on days with the same day-of-year as day t. The results in both panels are from CSDID (Callaway and Sant'Anna 2021). The unit of observation is city-month. All regressions control for home pollution.

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.[29] 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

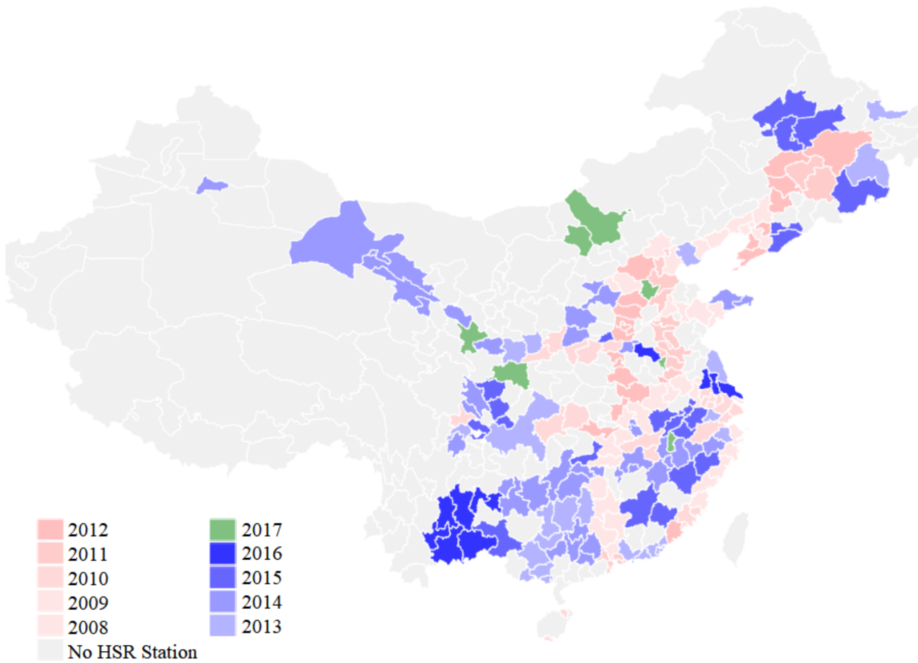
Table A1 presents summary statistics for key variables used in this analysis. There are a total of 330,819 city-day observations for $PM_{2.5}$ readings from 2013 to 2016. Some cities did not have $PM_{2.5}$ monitoring stations in the beginning of our sample, which explains $PM_{2.5}$'s lower number 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).

[29] 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

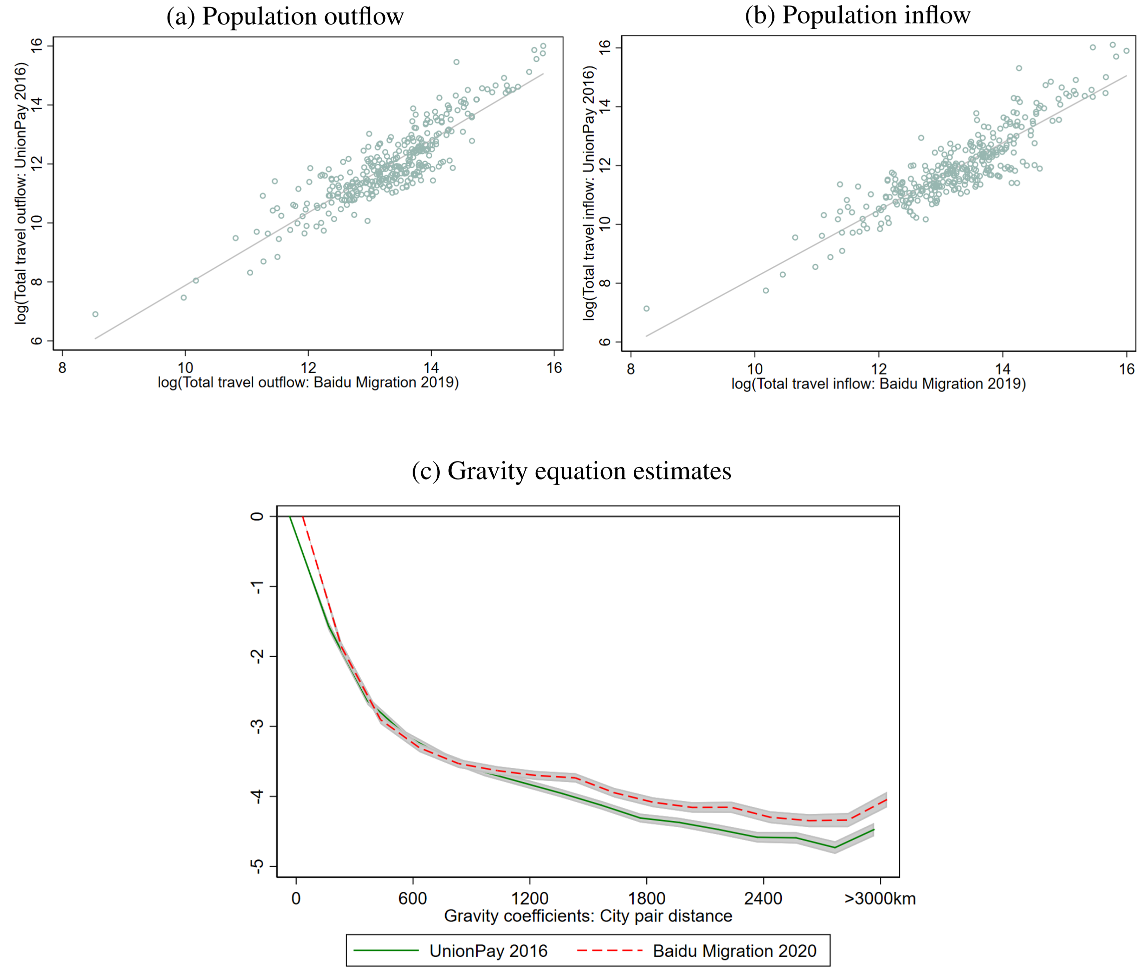
FIGURE A1. EXPANSION OF THE HSR NETWORK



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



Notes: Panel (a) plots the total number of outbound trips (in logarithm) 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 total number of arriving trips (in logarithm) 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 from the gravity equation below, based on city-pair bilateral trips from UnionPay data and the Baidu Migration Index, respectively:

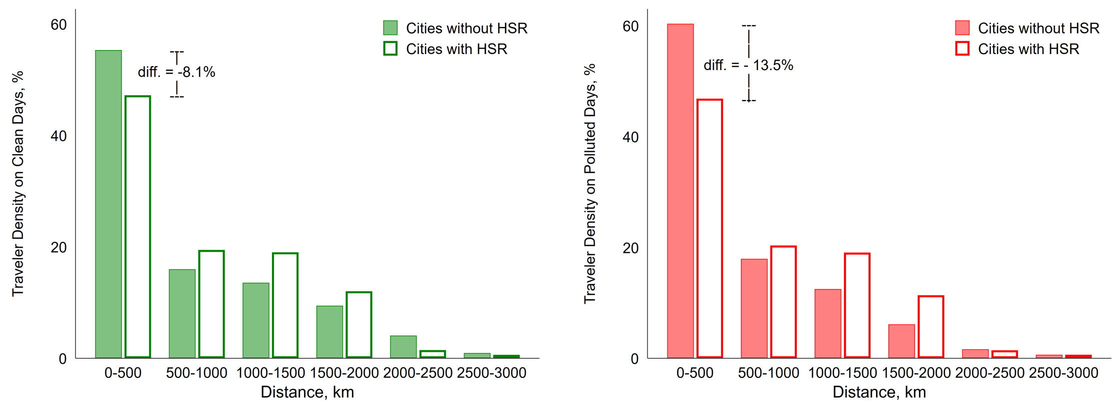
$$ln(N_{ij}) = \sum \beta_k I_k + \alpha_i + \gamma_j + \varepsilon_{ij},$$

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 (0-200 km is chosen as the base group) and α_i and γ_j are origin and destination city fixed effects.

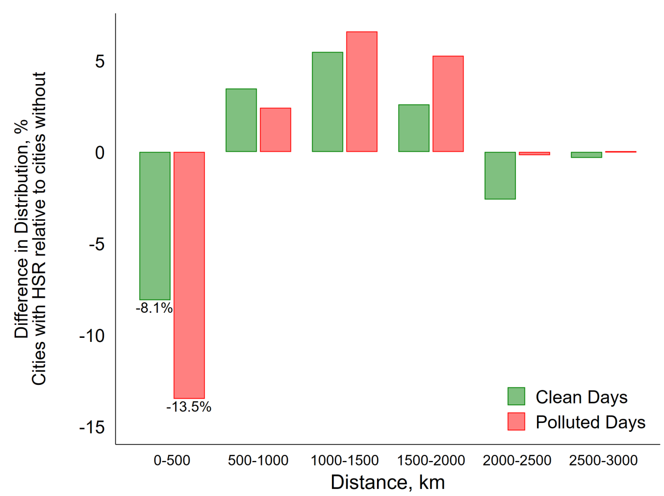
FIGURE A3. TRIP DISTANCE FOR CITY-DAYS DURING CLEAN AND POLLUTED DAYS

(a) TRIP DISTANCES DURING CLEAN DAYS

(b) TRIP DISTANCES DURING POLLUTED DAYS

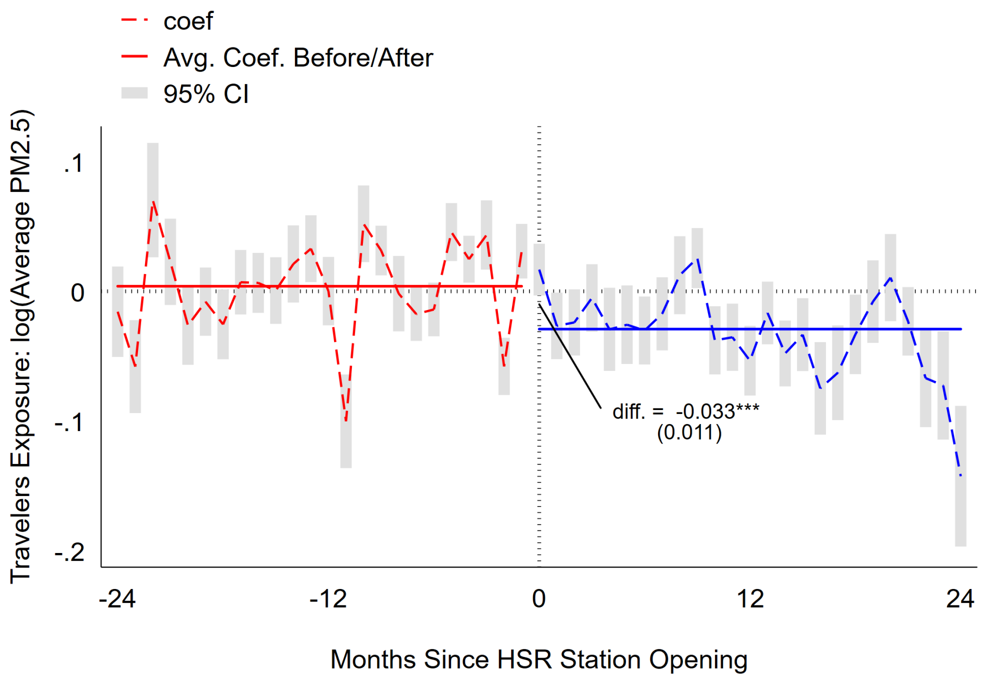


(c) DIFFERENCE IN TRIP DISTANCES



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 $\log(\text{PM}_{2.5})$



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

TABLE A2—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} (β)	-0.036	-0.034	-0.034	-0.032	-0.025
	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
Month-of-sample FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
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 2 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 HSR stations. The results in all columns are from CSDID (Callaway and Sant'Anna 2021). The unit of observation is city by month. We aggregate the data to the monthly level for CSDID instead of using daily data for computational tractability. 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 ($PM_{2.5} > 75\mu g/m^3$) at home city i for each month in all regressions. Standard errors are clustered at the city level.

TABLE A3— 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} (β)	-0.009	-0.009	-0.010
	(0.005)	(0.005)	(0.005)
Home Extreme (γ)	0.120	0.165	0.186
	(0.027)	(0.025)	(0.023)
Time-of-sample FE	Yes	Yes	Yes
City FE	Yes	Yes	Yes
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 2. The unit of observation is city by day in column (1), city by week in column (2), and city by month in column (3). HSR_{it} is an indicator variable for access to HSR stations. *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 but retain always-connected cities in our sample, which explains the difference in the number of observations compared to Column (1) from Table 2. The regressions are weighted by home cities' number of outbound travelers in period t. Standard errors are clustered at the city level.

TABLE A4— ROBUSTNESS TO TREATMENT OF MISSING POLLUTION VALUES

	Intensive margin: travelers' exposure to pollution extremes		
		$\text{PM}_{2.5} > 75\mu g/m^3$	
HSR_{it} (β)	-0.022	-0.034	-0.027
	(0.010)	(0.008)	(0.008)
Specification	drop missing home	no home control	dest interpolation
Month-of-sample FE	Yes	Yes	Yes
City FE	Yes	Yes	Yes
N	6,745	11,712	11,632

Notes: This table presents a set of sensitivity analyses of the HSR effect on travelers' exposure (the intensive margin) using equation (4) with different approaches for handling home and destination pollution data to that in Column (1) of Table 2. The dependent variable is a dummy variable indicating traveler's pollution exposure as defined in equation (1) to be over 75 $\mu g/m^3$. In Column (1), observations with missing home pollution measures are dropped. Column (2) excludes the control for home pollution. Column (3) employs an alternative method to address missing pollution values within $TP_{i,t}$. Specifically, missing values in $P_{j,t}$ (destination pollution on a specific day) are filled by interpolating using the pollution level observed on the same day the following year. HSR_{it} is an indicator variable for access to HSR stations. The results for all three columns are from CSDID (Callaway and Sant'Anna 2021). The unit of observation is city by month. Month-of-sample FEs and city FEs are included in all regressions. Standard errors are clustered at the city level.

TABLE A5—EXTENSIVE MARGIN USING SYNTHETIC CONTROL/DID APPROACH

	(1)	(2)	(3)
	Extensive margin: share of residents who travel		
HSR_{it} (β)	0.016	0.008	0.013
	(0.006)	(0.004)	(0.007)
Method	Synthetic Control	Synthetic DID	Synthetic Control
Specification	Arkhangelsky et al. (2021)		Ben-Michael et al. (2022)
N	11,664	11,664	11,664

Notes: This table revisits the effect of HSR on traveler shares (the extensive margin) using the synthetic control approach instead of CSDID as reported in Column (3) of Table 2. The specification in this table is the same as Column (3) of Table 2. Columns (1) and (2) report the unweighted ATT across post-treatment periods from synthetic control and synthetic DID estimations based on Arkhangelsky et al. (2021). Column (3) reports ATT weighted by the average number of residents in each city from synthetic control estimation based on Ben-Michael et al. (2022). The unit of observation is city by month. Month-of-sample FEs and city FEs are included in all regressions. Standard errors are clustered at the city level.

TABLE A6— THE EFFECT OF HSR USING SUB-SAMPLES

	(1)	(2)	(3)
	Excluding 5% days with most/least travels	Excluding 10% days with most/least travels	Excluding holidays and 40-day travel rush
	Panel (a): Intensive Margin		
HSR_{it} (β)	-0.032	-0.034	-0.032
	(0.008)	(0.009)	(0.009)
N	11,632	11,615	11,143
	Panel (b): Extensive Margin		
HSR_{it} (β)	0.007	0.006	0.008
	(0.003)	(0.003)	(0.003)
N	11,632	11,620	11,146

Notes: This table examines robustness of the intensive and extensive margins fo the HSR effect as reported in Table 2 to using different sub-samples. Panels (a) and (b) replicate Column (1) and (3) of Table 2 for the intensive and extensive margin, respectively. Column (1) (Column (2)) excludes each city's 5% (10%) of days with the most travels and 5% (10%) of days with the least travels. Column (3) drops holidays and the 40-day travel rush period surrounding the Chinese Spring Festival. The specifications for each regression are the same as the corresponding specifications in Tables 2 and the estimation is done using CSDID. The unit of observation is city by month. Month-of-sample FEs and city FEs are included in all regressions. Standard errors are clustered at the city level.

TABLE A7—HETEROGENEITY OF HSR EFFECTS WITH RESPECT TO HOME EXPOSURE

Panel (a): Intensive margin – travelers' likelihood of experiencing pollution extremes ($PM_{2.5} > 75\mu g/m^3$)			
	(1)	(2)	(3)
HSR_{it} (β)	0.009	Absorbed	0.006
	(0.008)	-	(0.006)
Home Extreme	0.189	0.189	0.239
	(0.012)	(0.012)	(0.021)
Home Extreme $\times$ HSR_{it}	-0.081	-0.081	-0.065
	(0.023)	(0.023)	(0.014)
Sample	Daily	Daily	Monthly
N	330,801	330,801	11,054
R^2	0.81	0.81	0.94

Panel (b): Extensive margin – share of outbound trips			
	(1)	(2)	(3)
HSR_{it} (β)	0.0127	Absorbed	0.0129
	(0.0080)	-	(0.0077)
Home Extreme	-0.0004	0.0002	-0.0044
	(0.0015)	(0.0013)	(0.0067)
Home Extreme $\times$ HSR_{it}	0.0009	0.0001	0.0058
	(0.0018)	(0.0016)	(0.0059)
Sample	Daily	Daily	Monthly
N	330,801	330,801	11,054
R^2	0.78	0.79	0.83

Notes: This table examines the heterogeneity of HSR impacts with respect to home exposure, the triple-difference variant of Table 2. See Table 2 for the definition of the dependent variables and regressors. All regressions are done using TWFE and weighted by home cities' number of outbound travelers in period t in Panel (a) and residents in period t in Panel (b). In both panels, Column (1) controls for day-of-sample FEs and city FEs; Column (2) further allows for the city-specific main effect of HSR (β_{2i}) which absorbs β; Column (3) controls for month-of-sample and city FEs. Standard errors are clustered at the city level.