

DISCUSSION PAPER SERIES

IZA DP No. 14900

**Green Infrastructure and Air Pollution:
Evidence from Highways Connecting Two
Megacities in China**

Bo Yu
Trang My Tran
Wang-Sheng Lee

NOVEMBER 2021

DISCUSSION PAPER SERIES

IZA DP No. 14900

Green Infrastructure and Air Pollution: Evidence from Highways Connecting Two Megacities in China

Bo Yu

City University of Macau

Trang My Tran

Monash University

Wang-Sheng Lee

Monash University and IZA

NOVEMBER 2021

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Green Infrastructure and Air Pollution: Evidence from Highways Connecting Two Megacities in China*

Following market liberalisation, the vehicle population in China has increased dramatically over the past few decades. This paper examines the causal impact of the opening of a heavily used high speed rail line connecting two megacities in China in 2015, Chengdu and Chongqing, on air pollution. We use high-frequency and high spatial resolution data to track pollution along major highways linking the two cities. Our approach involves the use of an augmented regression discontinuity in time approach applied on data that have been through a meteorological normalisation process. This deweathering process involves applying machine learning techniques to account for change in meteorology in air quality time series data. Our estimates show that air pollution is reduced by 7.6% along the main affected highway. We simultaneously find increased levels of ozone pollution which is likely due to the reduction in nitrogen dioxide levels that occurred. These findings are supported using a difference-in-difference approach.

JEL Classification: L92, O18, Q53, Q54, R41

Keywords: air pollution, China, green infrastructure, high-speed railway, regression discontinuity, machine learning

Corresponding author:

Trang My Tran
Department of Economics
School of Business
Monash University
47500 Subang Jaya
Selangor
Malaysia
E-mail: mytrang.tran@monash.edu

* We thank Shihe Fu, Jinhua Li, Shuddha Rafiq, Russell Smyth, Yanrui Wu for insightful comments as well as conference and workshop participants at China Economists Society 2021 conference, Central China Normal University, Peking University, Nanyang Technological University and Monash University. We also thank Tuan Van Vu for providing his weather normalisation code and his assistance regarding its use. This paper was previously circulated under the title "Does High-Speed Rail Improve Air Quality alongside Intercity Highways?". All authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1 Introduction

High speed rail (HSR) has often been argued to be a greener alternative to car and air travel. It is regarded as one of the most environmentally friendly forms of transportation as it is energy-efficient and has low emissions. HSR has gained favour in many countries as an intercity mode of transport (e.g., France, Germany, Italy, Japan) and many researchers regard HSR as a convenient substitute for intercity transportation modes (e.g., [Givoni and Dobruszkes, 2013](#)). In the intercity passenger transport market in contemporary China, HSR has emerged as an affordable and popular mode of travel. Although many aspects of HSR have been extensively examined in the literature, an understanding of the impact of HSR on intercity travel behavior changes and the resulting effect on emissions from highways and aviation that can be replaced by lower carbon emissions from HSR systems remains limited. Emissions from HSR are less than one-third of those from highways and much less than those from aviation ([Wee et al., 2005](#)).

This paper examines whether HSR can improve air quality by diverting marginal automobile travellers away from their vehicles. The substitution of HSR for road in intercity travel has important implications for air pollution in China. For example, exploring the idea that areas with greater improvements in railway services should also experience a more favourable evolution of air quality, [Lalive et al. \(2018\)](#) found that expanding regional rail in Germany caused a reduction in carbon monoxide (CO) and nitrogen oxides (NO_x) emissions. Of particular concern is the fact that China's vehicle population has increased dramatically over the past few decades as a result of market liberalisation, rapid social-economic development and urbanisation. The number of motor vehicles in 2017 was over 209 million, more than 13 times the number in 2000 (China Statistical Yearbook, 2000–2017). As motor vehicles are significant contributors of air pollution and have an important influence on air quality, there has been increasing alarm regarding the amount of on-road vehicular emissions generated. [Lang et al. \(2014\)](#) report a strong correlation between vehicular emissions and GDP at the province level. Severe ambient air pollu-

tion problems in many large cities of China have led to huge pressure to accelerate the progress on reducing vehicle emissions.

Utilisation of HSR services creates two opposing forces that influence air quality. On the one hand, increasing usage of HSR by the general population will increase the frequency of HSR services and decrease waiting times (the so-called Mohring Effect). An increasing substitution of HSR for car travel will eventually result in a reduction of air pollution. On the other hand, a significant investment in transportation infrastructure like HSR might lead to an increase in travel demand. If people travel more, this will generate more car trips to HSR stations, thereby leading to a traffic creation effect and generating more air pollution (Li et al., 2019). Taken together, the net effects of HSR utilisation on air pollution are therefore ambiguous.

This paper examines the environmental impact of the opening of a HSR line connecting two megacities in China in 2015: the Chengdu-Chongqing HSR line. The opening of the line reduced travel time between the two cities considerably, and was an attractive option for many intercity travellers. The Chengdu-Chongqing HSR line (opened in 2015), alongside the Beijing-Tianjin HSR line (opened in 2010) and the Beijing-Shanghai HSR line (opened in 2014), are some of the most heavily used HSR lines in China. Based on data from the Sichuan Statistical Yearbooks, Figure A1 shows that Chengdu saw a general decline in passenger loads for highway transportation while there was an increase in railway ridership (a combination of traditional rail and HSR) after 2015.¹ For commuting between Chengdu and Chongqing, following the opening of the Chengdu-Chongqing HSR line, 98.1% of the respondents in a survey indicated that HSR became their primary mode for intercity travel, with the percentage using private vehicles sharply dropping

¹Chengdu, with a population over 10 million, is the largest city in Western China which is mostly surrounded by mountains. It is the only large city in the huge Sichuan Province and is an attractive destination for many of the province's over 80 million people seeking work. Chongqing is one of China's most densely populated cities. Its huge infrastructure and industrial investment have made Chongqing grow to become one of the largest cities in China in the last five years. Chengdu and Chongqing are currently ranked as the sixth and seventh largest cities in China respectively. See: <https://www.chinahighlights.com/travelguide/top-large-cities.htm>.

from 23.4% before the opening of the line to 3.6% after the opening of the line. Furthermore, overall travel demand had increased by 60% (Ren et al., 2019). In this paper, we focus on the substitution effect of HSR for car travel and examine how the introduction of the HSR line between Chengdu and Chongqing affected air quality along the highways linking the two megacities.

Much of the prior literature has focused on the local effect of inner-city transportation on pollution (Chen and Whalley, 2012; Gendron-Carrier et al., 2021; Zheng et al., 2019). For example, Chen and Whalley (2012) present evidence of improving air quality in a local geographical vicinity using the discontinuity of subway ridership in Taipei on the day of subway opening. They show that the Taipei Metro opening in 1996 led to between a 5 and 10% reduction in CO but had no effect on ground-based ozone (O₃) pollutant concentration levels. Similar studies regarding the effects of urban transit rail on air pollution were also conducted in other megacities. Li et al. (2020) assess how subway line openings affect air pollution in Beijing by using staggered construction of subways over nine years. Their most conservative estimates suggest that a one standard deviation increase in subway density decreased a composite air quality index by two percent. In a similar fashion, by exploiting the staggered construction dates of subway lines, Zheng et al. (2019) show that the surrounding areas close to newly opened subway lines experienced a 18.1% reduction in CO in Changsha after one year.

Vehicle emissions have been traditionally measured using models such as COPERT, a European road transport emission inventory calculation model which serves as the European Union's standard vehicle emissions calculator. It calculates emissions from energy consumption for a specific place based on its vehicle population, mileage, speed and other weather data. Based on emission factors, fuel quality and annual vehicle kilometres travelled, Lang et al. (2014) applied the COPERT model to the Chinese context and estimated the annual vehicular emissions of different municipalities and provinces in China. Our work is an alternative approach to measure vehicle emissions and has the

advantage of allowing us to focus on a targeted location and at a higher frequency. We use gridded high-frequency satellite data to pinpoint the exact geographic location we are interested in for capturing pollution generated by vehicle emissions. By examining how air pollution along affected highways changes as a result of the opening of the Chengdu-Chongqing HSR line, we are directly measuring the impact of the operation of HSR on diverting marginal automobile travellers away from their vehicles.

To accurately measure air pollution along affected highways linking the two megacities, we identify the possible car travel routes using the optimal route planning function of Gaode Map. We focus on the top two highway routes used by drivers and collect air quality monitoring data sets that track the entire highway and which are located within 15 km of the affected highways (See Figure 1).

In order to estimate the causal effect of the opening of the Chengdu-Chongqing HSR line on air pollution, we use the opening date as a quasi-experiment and employ two complementary empirical strategies. First, we utilise a regression discontinuity in time (RDiT) model. The RDiT model is an adaptation of the more well-known regression discontinuity (RD) framework. Recent empirical work in several economic fields, particularly environmental and energy economics, has utilised the RDiT in applications in which time is regarded as the running variable and treatment is a specific time threshold (Hausman and Rapson, 2018). Our application of the RDiT approach in this paper follows the latest recommendations in the literature. We are careful to account for meteorological variation in the data using modern statistical methods involving machine learning techniques. We also use an augmented two-step RDiT approach that allows for better control of weather and seasonality factors. We provide more details on the machine-learning approach we use for the meteorological normalisation process to “deweather” the data and the augmented RDiT approach that we use later in the paper.

In contrast to the transportation literature that focuses on major cities that have been connected to HSR multiple times, this study focuses on an isolated HSR line involving

two of the largest cities in China. An advantage of such a focus is that it is more clear-cut to develop an appropriate causal model that attempts to accurately capture the specific local effects of the quasi-experiment. This helps with the credibility of our identification strategy.

Second, to supplement our main findings from the RDiT approach, we use a difference-in-difference (DID) approach that relies on a different source of identification. This approach involves the use of a control group that comprises random road segments that are unaffected by the opening of the HSR line of interest. While the RDiT approach relies on a discontinuity occurring at the time of the HSR opening for identification, the DID approach relies on a parallel trend assumption between the treatment and control groups. In our context, we expect the levels of air pollution in the primary expressways that drivers use to commute between Chengdu and Chongqing and other more distant expressways that drivers do not use to commute between Chengdu and Chongqing to have similar trends.

The closest paper to our work in the green transportation literature that studies air pollution along intercity highways is [Guo et al. \(2020\)](#). They employ a DID approach to estimate the effect of 23 HSR lines newly opened in 2015 and 2016 using air quality monitoring data from sparsely distributed monitoring stations around affected highways. They show that being connected to HSR reduces CO emissions from petrol cars near affected highways by 4.3% relative to the mean. However, they find the opening of HSR does not have statistically significant effects on ozone or diesel emissions such as atmospheric particulate matter (PM) that have a diameter of less than 2.5 micrometers (PM_{2.5}). We discuss in detail how our work differs from [Guo et al. \(2020\)](#)'s in Section 4 of the paper.

We find that the opening of the Chengdu-Chongqing HSR line led to a 7.6% reduction in overall air pollution concentration along the main affected Cheng-Sui-Yu (CSY) highway. In terms of heterogeneous effects by type of pollutant, we find that the decline in diesel-based pollutants after the HSR opening is larger along the CSY highway than

its less favoured alternative route, providing suggestive evidence that there was a larger decrease in intercity vehicle usage for the most affected highway. These findings are robust to multiple robustness checks such as alternative bandwidths and a "donut test" which excludes possibly sensitive observations around the cut-off day. Furthermore, increasing industrial activity which can lead to higher air pollution along the affected major highways does not explain our negative estimated coefficients.

At the same time, an unintended consequence we highlight is that we also find an increase in O₃ levels (a secondary pollutant) as a result of the opening of the Chengdu-Chongqing HSR line. This suggests that the opening of a new HSR line may not effectively reduce all types of air pollution. It highlights the complexity of using green infrastructure as a tool in air pollution regulation.

Overall, our contribution to the literature is the following. First, our high-frequency and high spatial resolution data allow us to examine a spectrum of pollutants that track the entire length of the affected highways over short time periods. Our approach to measuring vehicle emissions along highways is a useful alternative to traditional approaches that are based on collecting data on emission factors, fuel quality and annual vehicle kilometres travelled such as COPERT. Second, we use a state-of-the-art machine learning approach to account for meteorological variation and seasonality in the pollution data and adopt the latest methodological developments in the field to model the causal effects of the opening of a HSR line on air pollution. The former allows us to minimise bias in our estimated treatment effects as the functional form of the interactions between weather and air pollution are hard to specify correctly. Together, this allows us to establish how green infrastructure can explicitly affect local air quality alongside high density travel routes.

The remainder of this paper is organised as follows. Section 2 provides background on HSR in China. Section 3 describes the data. Section 4 provides a detailed discussion of the methodology used. Section 5 provides the results. Finally, Section 6 concludes.

2 Background

2.1 High-Speed Rail in China

HSR in China has developed rapidly for the past two decades with substantial investments from the Chinese government. From 2007 to 2017, China's total investments in HSR infrastructure exceeded 1,000 billion USD. China's 13th Five-Year Plan (2016-2020) had set a target to build 30,000 km of HSR connecting 80% of the country's major cities. By the end of 2020, China's HSR network had a total length of 37,900 km.² China currently has the largest HSR system in the world with 2 billion HSR trips taken per year.

Road congestion and traffic emissions are major contributors to air pollution worldwide. With enough people turning to greener modes of transportation such as HSR, vehicle usage can be reduced which would result in a corresponding improvement in air quality.³ National Emissions Inventories show that motor vehicles emissions in some large cities represent an important source of harmful ambient air pollution. In the U.S., CO and NO_x emissions accounted for about 34% and 38% of total emissions respectively in 2011.⁴ In China, car vehicle emissions have been a main source of air pollution, contributing to more than 60% of CO, around 40% of NO_x and at least 20% of PM_{2.5} in urban areas (Zheng et al., 2019).

HSR travel is considered to be a competitive substitute for aviation within a 1,000 km radius (Wang et al., 2017). It is particularly competitive when the distance is shorter. In fact, HSR is getting immensely popular in China and gradually replacing other modes of long-distance travel options such as airplanes, buses and private vehicles. As of 2020, the cumulative number of HSR passengers annually exceeded 2 billion. This soaring demand

²http://english.www.gov.cn/archive/statistics/202101/10/content_WS5ffa36f3c6d0f725769438ad.html.

³Chester and Horvath (2010) document that HSR lowered consumption of energy and the emission of greenhouse gases per trip. However, when factoring in how power for HSR is generated, they report that more SO₂ emissions were created given the sources used to generate electricity. This is not a major issue in our study because Sichuan province mainly uses hydropower.

⁴Environmental Protection Agency (EPA) (2015). 'Carbon Monoxide.' See <https://www.epa.gov/airquality/carbonmonoxide/>.

for HSR in China has led to a massive construction of the HSR network, which accounts for more than 60% of total HSR mileage in the world.

2.2 Background on the Chengdu-Chongqing HSR Line

The Chengdu-Chongqing HSR line was first planned in 2008 under a revised version of the four vertical and four horizontal HSR construction network that was setup for sustaining China's future economic growth. The construction of Chengdu-Chongqing HSR project began on the 22nd March 2010 and finished on the 26th December 2015 with a total length of 308 km. The tentative opening date of the HSR line was only precisely determined prior to the actual date, thereby helping to make the opening date exogenous for the purpose of our analysis. Due to the fact that the more economically prosperous regions in China are mainly coastal and in the east, most capital cities in the central and eastern provinces were connected to HSR network as a priority. Due to overlapping nodes and multiple train routes, the HSR network often includes many key provincial capitals in more than one HSR line. This makes examining the effects of HSR opening in a major provincial capital along China's eastern coast potentially more challenging. However, the case is different for the western provinces. HSR was not fully developed and intensively built there until much later.

Until 2015, there were only 3 HSR lines in the west. The Chengdu-Leshan (CL) and Chengdu-Jiangyou HSR lines were opened in 2014, and the Chengdu-Chongqing line opened in December 2015. The fact that these HSR lines are relatively isolated from other major cities in China is an advantage for the purposes of our analysis as air pollution in these local areas is less likely to be confounded by other government policies and unobservable factors.

The alternative mode of travel from Chengdu to Chongqing is by motor vehicles, including intercity buses and private vehicles. Figure 1 depicts the top two optimal driving routes between Chengdu and Chongqing: the Cheng-Sui-Yu and Cheng-Yu high-

ways. Table 1 shows that the length of the Cheng-Sui-Yu (CSY) Highway is 322 km (i.e., Chengdu Centroid to Chongqing Centroid through the CSY highway) and it takes 4 hours to complete the route, while the Cheng-Yu highway (CY) is 344 km long and it takes 4 hours and 30 minutes to get to the same destination. Highway tolls for both highways are effectively equivalent. The two recommended routes by Gaode Map are the ones most preferred and used by drivers and therefore most affected by the opening of the Chengdu-Chongqing HSR line. The CSY highway was fully opened in 2007 and about 12 years after the CY highway. Road conditions for the former are better and allow for higher speeds, making CSY the preferred highway to use when a non-stop journey is to be made between the two cities.⁵

Our identifying assumption relies on the fact that utilisation of these highways and commuting patterns will be affected with the opening of the Chengdu-Chongqing HSR line. One concern for our identification strategy is the number of concurrent environmental policies that China has implemented over the years. These policies can make isolating the effect of opening the HSR line on air pollution difficult. For example, the State Council announced the 12th Five-Year Plan in 2011, which regulated industrial SO₂ and NO_x, and the 'Blue Sky Protection Campaign' to regulate air pollution in developed areas of China from the 11th June 2018 to 28th April, 2019. There was also a direct regulation targeting gasoline content (the so-called "Yellow Labelled Vehicle" restrictions) that focused on gasoline vehicles that fell below the National I emissions standard and diesel vehicles below the National III standard. These restrictions gradually evolved from 2013 onward and had the goal of removing high emission vehicles from roads across the nation by 2017. Sichuan province and Chongqing did not implement the gasoline content restrictions until late 2017, which falls outside our analysis window.⁶ Hence, none of those regulations targeting emissions during our sampling period directly relates to Chengdu

⁵Although we are not able to obtain the historical traffic data, the most recent 2020 Gaode Map traffic data confirm that the two indicated highways are the most popular and favourable routes, and that CSY is the more heavily used highway.

⁶See <http://www.sc.gov.cn/10462/10464/10797/2017/3/8/10416508.shtml> (in Chinese).

and Chongqing and does not cause problems for our identification strategy.

3 Data

3.1 Air Pollutants

Air pollution and meteorological data are obtained from the Chinese Air Quality Reanalysis (AQR) dataset jointly produced by the Chinese Academy of Sciences and China National Environmental Monitoring Centre. They assimilate data from more than 1,000 ground based pollution and weather monitoring stations.⁷ The spatial resolution of grid cell of this hourly dataset is $15 \times 15 \text{ km}^2$. This data provides surface concentration levels of six pollutants: $\text{PM}_{2.5}$, PM_{10} , SO_2 , NO_2 , CO and O_3 , as well as data on humidity, surface pressure, temperature, wind speed and wind direction.⁸ In addition to this data set, we collect other meteorological variables from NASA Modern-Era Retrospective Analysis for Research and Application Version 2 (MERRA-2). They record surface temperature across 42 atmospheric layers and ground precipitation with a spatial resolution of $55 \times 65 \text{ km}^2$ in three-hour intervals (the highest time frequency available to us). The vertical structure of temperature layers allows us to calculate values for thermal inversion, which is commonly regarded as a crucial variable affecting air pollution concentration in the environmental economics literature. Specifically, values of thermal inversion are calculated through temperature differences in the nearest two surface layers from the ground, following a common approach by [Chen et al. \(2017\)](#) and [Deschenes et al. \(2020\)](#).⁹ We merge the AQR data set with the nearest MERRA-2 variables and then vectorise precise coordinates of each AQR grid cell in the merged dataset. Taking advantage of the

⁷See <http://www.dx.doi.org/10.11922/sciencedb.00053> by [Kong et al. \(2021\)](#) for more details about the data.

⁸We calculated the wind direction angle via the arc-tangent function of the vertical and horizontal components of the wind vector provided by this dataset. Wind direction ranges between 1-360 degrees.

⁹We successfully replicated the trend of thermal inversion strength in [Chen et al. \(2017\)](#) using China's national counties list. Figure [A2](#) depicts the trend of thermal inversions. Although our trend is not exactly quantitatively the same as in [Chen et al. \(2017\)](#), our replicated trend is qualitatively similar.

high spatial resolution of the AQR grid cells, we are able to trace the nearest data points along each affected expressway. This allows us to collapse the multiple AQR grid cells into an average ‘expressway level’ measure of pollution for our analysis.¹⁰ Figure 1 displays the nearest AQR grid cells that track the three expressways of interest in this study.

There is heterogeneity in the composition of pollutants that different types of vehicles emit, which is important to note in our context. Petrol car emissions comprise mainly of CO (more than 80%), while NO₂ is a major pollutant of combustion diesel fuel engines and O₃ is an indirect pollutant from cars.¹¹ The amount of particulate matter that is generated is dependent on traffic congestion. Table 2 provides summary statistics of hourly air pollution measurements for the CSY highway, our main highway of interest. Columns (2) and (7) present the sample averages of pollutant concentrations and various weather indicators for the period one year before and one year after the opening of the Chengdu-Chongqing HSR line.¹² The descriptive statistics provide preliminary evidence that most of the direct automobile-related emissions are reduced after the Chengdu-Chongqing HSR line opened. For example, the mean concentration of CO reduced from 0.849 to 0.819 mg/m³ and mean NO₂ dropped from 21.71 to 21.15 µg/m³. Surprisingly, ozone emissions are found to have increased one year later, suggesting that not all types of air pollutants decline after the HSR opening (we discuss this inverse relationship and possible reasons for this finding in more detail in Section 5.3).

It is worth noting in Table 2 that the HSR opening did not lead to a sizeable change in the mean values of the meteorological variables. This lack of a change suggests a smooth transition in weather conditions after the treatment date. However, this smooth transition of weather around the cut-off would still not guarantee a clean identification if weather

¹⁰Our results remain quite robust to using either the collapsed or non-collapsed AQR cell data.

¹¹Ozone is not emitted directly from automobiles, but the unstable compound is formed in the atmosphere through a complex set of chemical reactions involving hydrocarbons, oxides of nitrogen, and sunlight.

¹²We choose one year as the baseline time window in order to allow the treatment effect to take place. This is similar to the time window used in the previous literature that examined the association between transportation infrastructure, investment and high-frequency air pollution (Chen and Whalley, 2012; Zheng et al., 2019)

events interact with air pollutants over time. Following recommendations in the literature, we address this issue by establishing predicted counterfactual observations using a machine-learning-based random forest algorithm.

3.2 Deweathered Air Pollution Data

Accurate measurement of the impact of the opening of the HSR line on air pollution requires a trend analysis of ambient air quality data. However, wind speed, wind direction, surface pressure, atmospheric temperature, humidity and thermal inversion can have large influences on pollutant concentrations (Carslaw and Ropkins, 2012). Such factors cause complications as the influence of such meteorological variables can be larger than the effect of the intervention that is the focus of attention. In other words, it is usually unknown if the variation of the pollution trend is driven by changes in meteorological conditions or changes in emissions (Grange et al., 2018).

The non-linear relationship between the concentration of contaminants and their emission dispersion is difficult to capture using traditional deterministic models (Grange et al., 2018). As the interactions between air pollutants and their influential factors can be complex, in such instances, a machine-learning algorithm can be a useful tool to help model the non-linearity of air pollution. In other words, to see the real benefits in air quality by different policy interventions, it is essential to correctly decouple the meteorological impact from ambient air quality data (Grange et al., 2018; Grange and Carslaw, 2019; Vu et al., 2019).

Grange et al. (2018) found that machine-learning-based techniques outperform traditional statistical and air quality models by reducing variance/bias and error in high-dimensional data sets. Their approach was to apply a decision tree machine learning method and to use this predictive model to predict counterfactual ambient atmospheric pollutant concentrations for their trend analysis. Specifically, their weather normalisation technique involves the use of a random forest (RF) algorithm that predicts differ-

ent air pollutants at a specific time by re-sampling from a predictor data set. A repeated sampling process and RF model predictions of individual cells can help establish the meteorological normalisation of a daily air quality data series. In their approach, every air pollutant concentration was predicted 1,000 times by the RF model. For each prediction, the predictive features were sampled without change and were randomly allocated to match the air pollutant concentration. Finally, the arithmetic mean of each predicted pollutant was calculated using the aggregated 1,000 predictions, representing the meteorologically normalised trend that would be used for further analysis.¹³

However, in their application, [Grange et al. \(2018\)](#) not only deweathered the data but also removed time trends from their normalised data. As a result, it is unable to investigate the seasonal variation in the trends of pollutants that correlate with economic cycles and vehicle traffic. For this reason, this paper adopts the meteorological normalisation procedure used in [Vu et al. \(2019\)](#).

Appendix Figure [A3](#) illustrates the construction process of a RF model that predicts counterfactual deweathered air pollutants. The algorithm first builds the model by splitting the sampling period of interest over 2013 - 2017 into a training data set (70% of the data) to train the RF model and a test data set (30% of the data) to test model performance. Once the RF model is built, following the algorithm used by [Vu et al. \(2019\)](#), this paper first sets original time variables and resamples weather data (wind speed, wind direction, temperature, and relative humidity, surface pressure, thermal inversions and precipitation) as a new input data set of predictor features. Specifically, these weather variables at a specific selected hour of a particular day used as predictor variables were generated by randomly selecting from the recorded AQR and MERRA-2 weather data over the longest period we have from 2013 to 2018 at that particular hour of different dates within a four-week period (that is, 2 weeks before and 2 weeks after that selected

¹³One caveat is that the random-forest algorithm works best with large data sets because using RF models on small data sets runs the risk of overfitting. This is not an issue in the case of this study as we use high-frequency hourly pollution data over a five-year sample.

date).

For example, for a particular date and time (e.g., 09:00, 15th March 2016), the new predictors are constructed only by using the randomly selected weather variables from the original data set within a four-week interval (i.e., at 09:00 on any date between the 1st and 29th March on any year during 2013 to 2018). A final new input data set over 2013 to 2017 is generated by automatically repeating the selection process for all time and dates within the period of interest and then fed to the RF model to construct the predicted pollutant concentration. This step is repeated 300 times.¹⁴ Finally, for each particular hour, day and year, these 300 predicted output values are averaged to calculate the final weather normalised concentration. This way, unlike in [Grange et al. \(2018\)](#), [Vu et al. \(2019\)](#) only normalise the weather conditions but not the seasonal variation in the pollution data.¹⁵

We present the summary statistics of the machine-learning predicted air pollution data (deweathered data) in Table 2. Both the weather normalised data and the raw pollution data show a similar pattern of change in pollution levels. However, the deweathered data show two distinct features. First, although the mean values of each pollutant remain at a similar level, the standard deviations in column (3) are much smaller. This suggests that air pollution trends are much less volatile after eliminating the confounding weather conditions. This is confirmed by closely examining air pollutant concentration along the CSY highway in Figure 2. The figure clearly highlights that the volatility of pollutant concentrations dropped dramatically following the deweathering process. As a result, the predicted counterfactual pollution trends are much smoother.¹⁶ Second, as shown in

¹⁴One can of course increase the number of predictions. However, similar to [Cole et al. \(2020\)](#) and [Shi et al. \(2021\)](#), we find that the predicted air pollutant concentrations have been stable at 300 times predictions through our tests.

¹⁵The code is available on Vu's personal website. One issue with the use of these packages and the associated machine-learning algorithm is that it requires a very heavy computational workload. In our paper, to generate 300 predictions in total, around 20,000 computing hours was required. This was to deweather 5 pollutants measured at an hourly level in 55 air pollution grid cells over a period of 5 years.

¹⁶Box plots in Figure A4 in the appendix further reveal that the RF prediction reduces the volatility of AQR pollution data on different quantiles across different hours in a day for all pollution grid cells.

columns (4) and (5), the recorded minimum values of raw NO₂, particulate matter and O₃ in the sample are zero. This can occur when extreme weather conditions confound pollution concentration levels so that these pollutants cannot be detected. By removing weather factors, we find that minimum values of the deweathered pollutants increase and the maximum values decrease. This process helps ensure that the actual pollution concentrations are in a reasonable range. Subsequently, the results in the regression models used later will be less likely driven by extreme weather conditions.

Altogether, consistent with previous studies ([Grange et al., 2018](#); [Grange and Carslaw, 2019](#); [Vu et al., 2019](#)), the descriptive statistics suggest a need to account for weather patterns and meteorological conditions. The use of machine-learning prediction techniques to deweather the data helps provide us a clearer understanding of the actual ambient pollution trends.

4 Methodology

Our approach to analysing the effects of HSR on pollution alongside intercity highways is based on using the opening of the HSR line as a quasi-experiment. Our main approach involves the use of an augmented RDIT approach paired with pollution data that have been appropriately deweathered. An alternative evaluation technique we use to supplement our main findings is a DID approach. We discuss our implementation of both these approaches in more detail in the subsections below.

4.1 Regression Discontinuity in Time

In order to help estimate the causal effects of a policy on air pollution, one empirical strategy is to compare pollution levels at peak hours before and after policy implementation, while controlling for day of the week and month of the year fixed effects, economic activity and meteorological conditions ([Chen and Whalley, 2012](#)). As there may be an en-

dogeneity problem regarding the introduction of the policy, one way around this issue is to focus exclusively on the short-run impact of the policy using a regression discontinuity design (Chen and Whalley, 2012; Gallego et al., 2013). We use the regression discontinuity in time (RDiT) estimator of Hausman and Rapson (2018) that is suitable in our context given the availability of high-frequency data.

In this section, we begin by using a standard RDiT approach of Chen and Whalley (2012), which is given by

$$\begin{aligned} Air_Pollutants_{i,h,t} = & \beta_0 + \beta_1 OpeningHSR_t + \beta_2 \mathbf{P}(\mathbf{t}) + \beta_3 \mathbf{P}(\mathbf{t}) \times OpeningHSR_t \\ & + \beta_4 X_{i,h,t} + Pollutant_i + \epsilon_{i,h,t}. \end{aligned} \quad (1)$$

$Air_Pollutants_{i,h,t}$ is the logarithm of pollutant concentration i (CO, NO₂, PM_{2.5}, PM₁₀ and O₃) in hour h of day t for affected highways based on the Gaode optimal route mapping system.¹⁷ $OpeningHSR_t$ is a dummy taking a value of 1 if HSR is operational and 0 for all the days prior to the opening date on the 26th December 2015. $\mathbf{P}(\mathbf{t})$ is a vector of a third-order polynomial time trends that controls for temporal variation in air pollution that may have occurred when HSR inauguration was absent (Chen and Whalley, 2012). The $X_{i,h,t}$ includes quartic terms and 1-hour lags in temperature, wind speed, wind direction, humidity and surface pressure in conjunction with 3-hourly and lagged thermal inversions and precipitation. It also includes hour and day of (Julian) year fixed effects and their interactions between hour by week fixed effects and day of Julian year fixed effects. In addition to these meteorological control variables and temporal fixed effects, we also include lagged air pollution concentration levels.¹⁸ We stack all the pollutant variables to assess the overall effect of HSR openings and include a $Pollutant_i$ dummy to control for heterogeneity in time invariant pollutant concentration levels, following Duflo

¹⁷We do not include SO₂ in the analysis because it is a minor vehicle pollutant. Power plants emit more than 60% of total industrial SO₂ and all industrial manufacturing firms including power plants emit more than 80% of total SO₂ emissions (Yu et al., 2020).

¹⁸An (unreported) estimate of 0.81 in the AR(1) process confirms a strong auto-correlation in air pollution concentration levels, suggesting the need to include a lagged dependent variable.

et al. (2013) and Li et al. (2020).

Davis (2008) suggests that seasonality can be hard to control for with less than two years of data. For estimating our RDiT model using Equation 1, we therefore choose to use a two-year pre-post time window (i.e., 4 years of data) relative to the HSR opening date for our analysis. Standard errors are clustered at five-week intervals to reflect the level of persistence of carbon monoxide in our sample (Chen and Whalley, 2012; Davis, 2008).

4.2 Augmented RDiT

Our preferred model for the analysis in this paper is an augmented RDiT model that addresses the over-fitting issue of the global polynomial time trend in the standard time series discontinuity setup. According to Hausman and Rapson (2018), the augmented two-step RDiT approach allows for better control of weather and seasonality factors.

The identification assumption in our RDiT model is that air pollution levels along the highway will change smoothly in the absence of the new HSR line opening, conditional on the various controls we include and allowing for flexible fixed effects. By focusing on pollution levels around the highway, the treatment effect we estimate reflects the extent to which HSR is used as a substitute for motor vehicle travel when travelling between Chengdu and Chongqing.

In the first stage of RDiT model, data are stacked in a similar style to Equation 1 and the model is estimated as follows:

$$Air_pollutants_{i,h,t} = \psi + \lambda_t + Holiday_t + Pollutant_i + \epsilon_{i,h,t}, \quad (2)$$

where $Holiday_t$ is a holiday dummy, given a value of 1 if the day of year is China official holidays or important festivals.¹⁹ λ_t indexes a set of time fixed effects including hour fixed

¹⁹Dates of Chinese traditional festivals such as the 7-day Spring Festival, 3-day Dragon Boat Festival, Mid-Autumn Festival and Tomb-Sweeping Day change across different years. As a robustness check repor-

effects (FEs) and day of (Julian) year fixed effects. The first stage model over a 4-year window (2014 Jan - 2017 Dec) allows for temporal fixed effects that capture seasonality in air pollution and unobserved business cycles that would have occurred before the opening of the Chengdu-Chongqing HSR line. For example, an operational Chengdu-Chongqing HSR line can intensify population flows during *Chunyun*, the largest annual internal migration in the world around January and February before the Chinese Spring Festival. If Chunyun-associated activities create more emissions, the lack of temporal fixed effects can create a downward bias in our estimates. If all the confounding factors are successfully accounted for in the first stage, estimation can be regarded as quasi-experimental in the next stage of the RDiT model.²⁰

In the second stage, we proceed to estimate a local linear regression by regressing residuals from the first stage on lagged air pollutants, the HSR opening date and time trend as a running variable. The local linear regression model estimates around both sides of the cut-off points would help avoid over-fitting issues that would result from higher-order global polynomials. For the local linear specifications, we use a uniform kernel function and one year of data for pre and post the opening date.^{21,22} We also follow [Hausman and Rapson \(2018\)](#)'s recommendation and bootstrap the standard errors using 500 repetitions in order to reflect the standard error from the first stage. The second-stage

ted in the Table 3, we show that the baseline estimates barely changed, confirming that the temporal FEs can well capture the change in trends of air pollution due to the surge in traffic during holidays in different years.

²⁰Given the augmented RDiT is based on a four-year time window in the first stage, a placebo exercise would require at least four years of air pollution data before 2016 to randomly assign a false treatment date. The sample period for our air pollution data limits us from conducting such a test. However, we conduct other sets of robustness checks in Section 5.5 to examine whether our main results still hold in different scenarios.

²¹Our estimates remain robust to using a triangular or epanechnikov kernel.

²²A narrower two-year window (one year each for pre- and post-periods) in our second stage is a common choice for RD analysis in the green transportation literature (e.g., in [Chen and Whalley \(2012\)](#)). Our time-series setting implies that the optimal bandwidth method of [Calonico et al. \(2014\)](#) or [Imbens and Kalyanaraman \(2012\)](#) is not applicable due to very limited cross-sectional variation.

model is presented as follows.

$$\begin{aligned} \text{Residual}_{i,h,t} = & \beta_0 + \beta_1 \text{OpeningHSR}_t + \beta_2 \text{Day}_t + \beta_3 \text{OpeningHSR}_t \times \text{Day}_t \\ & + \beta_4 \text{Air_Pollutant}_{i,h,t-1} + \epsilon_{i,h,t}. \end{aligned} \quad (3)$$

$\text{Residual}_{i,h,t}$ is the predicted residual from the first stage. Day_t is normalised to be zero on the day the HSR line opens, following [Anderson \(2014\)](#), i.e., $\text{Day}_t \in [-365, -364, \dots, -2, -1, 0, 1, 2, \dots, 364, 365]$.²³ Our coefficient of interest is $\hat{\beta}_1$ that captures the short-term HSR effect, and it will be unbiased if the RD assumption holds.

4.3 Difference-in-Difference

The augmented RDiT design examines the local effect of HSR immediately after its opening. As an additional check, we conduct a DID analysis to help supplement our main RDiT analysis.

Our DID approach uses control groups based on random road segments and is similar in fashion to the approach used in [Gu et al. \(2021\)](#). Identification in our DID model is based on assuming similar time trends in air quality for expressways closer to and further away from the HSR line. To operationalise the DID analysis, we choose AQR grid cells of some random road segments that are further away and which intercity drivers cannot use to directly commute to the CSY and CY highways. Therefore, they are expected to be unaffected by the opening of the Chengdu-Chongqing HSR line.

We restrict all control road segments to be within Sichuan and Chongqing provinces. These selection criteria help ensure that these AQR grid cell along the control road segments are more likely to share similar characteristics to the highways of interest. To im-

²³We take day rather than hour as the treatment cut-off because there were more than 20 HSR departure times each day between Chengdu and Chongqing. Commuters diverted to HSR can choose the schedule that suits them best within a day, so the exact treated hour is unclear. We therefore seek a higher unit of time (day) as the treatment. Our RD estimates (discussed later) remain robust to using different hours within the 26th December 2015 as a threshold. The same practice using day as a cut-off in a regression discontinuity analysis can be seen in [Anderson \(2014\)](#).

plement our DID analysis, we estimate the following equation.

$$Air_Pollutant_{c,s,t} = \beta_0 + \beta_1 OpeningHSR_t \times AQR_{c,s} + \beta_2 X_{c,s,t} + \lambda_t + \lambda_c + \epsilon_{c,s,t}, \quad (4)$$

where $Air_Pollutant_{c,s,t}$ is the natural logarithm of deweathered air pollution concentration recorded in AQR pollution grid cell c of road segment s in day t . $AQR_{c,s}$ takes on a value of one for all AQR observations belonging to the road segments along the CSY highway, and is zero for AQR observations c of segment s in the control road segments. $X_{c,s,t}$ is a set of a third-order polynomial weather controls to improve balancing between treated and control AQR grid cells. λ_c indexes the AQR grid cells fixed effects. λ_t represents the interactive year-by-day fixed effects, which captures shocks common to all AQR cells over the whole sample. Standard errors are adjusted by clustering on road segments to allow for autocorrelation over time and spatial correlation within each segment.

As previously mentioned, the closest paper to our work is [Guo et al. \(2020\)](#) who employ a DID approach to estimate the effect of 23 HSR lines newly opened in 2015 and 2016 on pollution around affected highways. Our work differs from [Guo et al. \(2020\)](#)'s in several respects. First, our high-frequency and high spatial resolution data that span 2013-2017 allow us to trace the full spectrum of the change in air pollution along the affected highways of interest. Our longer time series data allow us to trace the trends of air pollution along highways over a few years, enabling us to use the RDiT approach to control for seasonal trend change while detecting any resulting discontinuities that arise. [Guo et al. \(2020\)](#) focus on a wider geographic scope and use sparse data records from air pollution monitoring stations that are more spread out. This limits their analysis to having only a few monitoring stations close to highways, sometimes just one or two. In contrast, we use data from 55 AQR grid cells to closely track changes in air pollution levels along the entire length of three highways.

In addition, our application of a DID model differs in the setup. In their DID analysis,

Guo et al. (2020) define those stations within 10 km of highways affected by the HSR lines opened in 2015 and 2016 as treatment stations, and their control group to be HSR stations that are within 10 km of highways affected by the HSR lines which opened before 2015. Rather than relying on an assumption that air quality in highways close to HSR stations that opened before and after 2015 have the same time trend for identification, we use a distance-based DID approach. Our control group is selected by locating random road segments and their associated AQR grid cells at least 30 km away from the HSR line. In other words, our DID identification is based on assuming similar time trends in air quality for expressways closer and further away from the HSR line. This approach therefore helps us estimate a local average treatment effect that is related to the local effect that we estimate using the RDiT model and helps to serve as a robustness check.

5 Empirical Results

5.1 Estimates Based on Standard RDiT

Table A1 presents the results from Equation 1 for results over the period 2014-2017. The four-year window should allow enough time to control for seasonality factors. *OpeningHSR* is our variable of interest. Columns (1) - (6) show the estimates using raw air pollution concentrations in logs by controlling for temporary fixed effects and pollutant dummies (for stacked pollutants only). The estimated coefficients suggest a slight positive but statistically insignificant relationship between HSR opening and most pollutant concentrations. However, the true statistical relationship between weather and pollutants is hard to correctly specify. We therefore use a RF based approach to establish deweathered counterfactuals that isolate weather effects.

We repeat the same exercise using deweathered pollution concentration levels in columns (7) - (12). All the coefficients are either insignificant or indistinguishable from zero, suggesting opening of HSR had no association with air pollutants. However, there

are two limitations in using traditional time-series RD in our case. First, a common practice that adds third-order polynomial time trends may be subject to over-fitting that biases our estimator (Hausman and Rapson, 2018). Second, a long-time window is generally better at capturing seasonal confounders. However, an issue is that a long time window can increase the likelihood of the presence of unobserved variables. To address these two issues, we utilise the augmented version of the two-step RDiT model whose results are discussed in the next subsection.

5.2 Main Results: Augmented RDiT for the Main Highway

Table 3 reports our baseline results. It shows regression estimates using the stacked pollutant variable to examine the comprehensive pollution impacts using a one year pre-post window. Column (1) shows the baseline results using non-machine-learning processed AQR pollution data, where weather controls have been included. These results suggest that opening of the HSR line caused a 11% reduction in air pollution concentration. When we further include holiday and day of (Julian) year (day = 1, 2, ... , 364, 365) fixed effects in Columns (2) and (3) of Table 3, the results suggest that omitting these fixed effects can lead to an underestimate of the effect.

The inclusion of weather controls can mitigate the weather effects on air pollution to some extent. However, weather has a complex interaction with air pollution in nature. Hence, a false assumption made on the interaction of meteorological conditions on air pollution can lead to bias in estimating the effect of the HSR line opening. Columns (4) to (6) present similar results to columns (1) to (3) where now deweathered pollution data are used instead. The resulting estimates are half the size as compared to the results based on using raw pollution data. The values of pollution concentration based on the machine learning process is less dispersed and volatile after the deweathering process. Therefore, less of a jump is seen around both sides of the cut-off based on the deweathered data. Our preferred estimates show that air pollution is reduced by 7.6% along the main affected

highway (column 6).²⁴

5.3 Heterogeneous Effects

Pollutant heterogeneity — Instead of focusing on stacked pollution, we now focus on assessing the importance of each individual element. Using raw data, Figure 3 plots residuals from regressions of air pollutants on a set of fixed effects by estimating Equation 2, adjusting for seasonality and shocks common to all pollution grid cells for the CSY highway. Each point in the figure represents residuals collapsed into two-week bins and the dashed lines indicate the beginning of Chengdu-Chongqing HSR. We find a decline in pollution concentration levels for CO and particulate matter (PM_{2.5} and PM₁₀), while we see a slight decline in NO₂ and increase in O₃ emissions along the CSY highway after the HSR line opening. Figure 4 suggests a similar pattern using deweathered data, where the only difference is that the residuals in this case from using the deweathered data are less dispersed.

Table 4 presents formal regression estimates of the results seen in Figure 3 and 4 using an augmented RDiT model. Each column reports the result of a separate regression. Using raw pollutant data, columns (1) - (5) show that there is a 14% and 11% reduction in CO and NO₂ respectively in the area along the affected CSY highway, following the opening of Chengdu-Chongqing HSR line. We also observe an approximate 30% reduction in particulate matter. These values are much larger than the estimates in the current green infrastructure literature (Chen and Whalley, 2012; Guo et al., 2020; Lalive et al., 2018). A possible reason is that there is a misspecification in depicting the relationship between meteorological variations and air pollution, leading to bias in our estimates.

We therefore utilise deweathered data that are orthogonal to any functional form

²⁴Denoting the coefficient of lagged air pollutant as α , the total HSR effect on air pollution is a sum of short-term effects dissipating over an infinite time horizon following Hausman and Rapson (2018), such that the total effect is $\hat{\beta}_1 + \hat{\beta}_1 \cdot \hat{\alpha} + \hat{\beta}_1 \cdot \hat{\alpha}^2 + \hat{\beta}_1 \cdot \hat{\alpha}^3 + \dots + \hat{\beta}_1 \cdot \hat{\alpha}^n = \hat{\beta}_1 / (1 - \hat{\alpha})$ when $\hat{\alpha} \in (0, 1)$. Hence, our baseline estimate reported in Column (6) of Table 3 suggests that air pollution over the long-term is reduced by 8.71% ($= 7.6\% / (1 - 0.128)$).

assumptions between weather and air pollution. Columns (6) - (10) show that the petroleum-based emission, CO, along the CSY highway decreases by 8% after the HSR line opening. This is in line with the range of estimates in the existing literature that recorded reductions in CO to be between 2% and 14%, suggesting that our estimated effects are plausible. We also find a reduction in diesel-based pollutants such as particulate matter (10% for PM_{2.5} and 6% for PM₁₀) and a 3% decline in NO₂. The reduction in CO suggests a reduction in private vehicle usage, while a reduction in particulate matter and NO₂ suggests a diversion of transportation mode away from intercity passenger buses. For example, Figure A1 displays a decline in intercity highway usage for Chengdu drivers and increase in railway services. Although we cannot disentangle the exact composition of the use of traditional railway and HSR, this figure helps provide indirect evidence that passengers are using HSR as a substitute for intercity bus trips, which can lead to a decline in diesel consumption. Together with the survey results from Ren et al. (2019), these results present strong evidence that passengers prefer the faster and greener HSR as a mode for intercity transport after its opening that lead to lower levels of air pollution.

Despite a reduction in primary emissions from vehicles, the regression estimate shows a 1.3% increase in O₃, which is not directly emitted from motor vehicles. However, this increase in ozone emissions in our results is not implausible given the complex interactions and chemical processes involving O₃. This is because O₃ is a secondary pollutant and chemically interacts with NO and NO₂.²⁵ The chemical interactions between O₃ and other pollutants and weather conditions are complex, especially when gasoline pollutants such as volatile organic compounds (VOCs) are involved. Lower VOCs and NO_x (NO₂+NO) levels will lead to lower O₃ levels. The change in NO_x levels can result in variation in O₃ levels that depend on the relative dominance of VOCs and NO_x. An inverse

²⁵NO is another automobile pollutant. Unfortunately, the AQR data did not record this pollutant. However, NO usually strongly correlates with NO₂ implying that we can infer a trend change in NO using NO₂. A simple relation between O₃ and nitrogen oxide emissions is: $NO + O_3 \rightarrow NO_2 + O_2$.

relationship between O_3 and NO_2 is sometimes seen in the science and environmental economics literature (Sager, 2019; Shi et al., 2021).

The increase in ozone pollution has multiple detrimental effects such as productivity losses and higher health costs. For example, Wang et al. (2020) show that a one standard deviation increase in ozone concentration can lead to a decrease in productivity of workers in five top express courier companies in China by 37.9%, amounting to a welfare cost at almost 300 million USD in China in 2019. Emerging evidence has shown that both short-term and long-term exposures to ozone are associated with increased mortality due to respiratory and cardiovascular diseases (Zhang et al., 2019).

Alternative affected highway — Panel A of Table 5 shows the estimation results for the CY highway. This is an alternative highway route from Chengdu to Chongqing displayed in Figure 1. The road conditions in the CY highway are not as good as in the CSY highway because this highway was built much earlier. At the time it was built, it was not designed to accommodate the high density of road traffic that arose a decade later. However, some drivers might still choose to drive on this highway depending on their intended stopovers along the way.²⁶

The estimated results in Panel A of Table 5 show that the reduction in CO levels is similar to that found for the CSY highway. This implies that the opening of the HSR line exerts similar effects for drivers along both highways. However, the magnitude of reduction in diesel-related emissions, particulate matter and NO_2 , is smaller. This would suggest a smaller decline in intercity buses using the CY highway. This corresponds to

²⁶To the extent that HSR facilitates intercity travel, this should generate more vehicle trips between intra-city locations and HSR stations. This could lead to more intra-city pollution. However, new subway lines were being built in both Chengdu and Chongqing over our sample period, connecting different parts of city to the HSR stations. This could lead to a reduction in urban air pollution (Gendron-Carrier et al., 2021; Li et al., 2019). For example, Chengdu Metro Line 3 opened on the 31st July 2016 and Chengdu Metro Line 4 opened on the 1st January 2016. In the same year, Chongqing opened Chongqing Metro Line 3 on the 28th 2016. Chengdu Metro Line 3 and 4 took around 200,000 passengers on the first day of its opening and passenger ridership for both subways towards 2020 were more than 700,000 each day (see an example from the official website of Chengdu Rail Transit at http://www.chengdurail.com/sw_detail/4096.html). These subways carry millions of passengers each day and can offset the increasing air pollution resulting from increasing travel between the HSR station and the urban centre. Therefore, the net effect on intra-city air pollution is inconclusive and ambiguous.

the results of Table 4 that the opening of the HSR line results presented a greater shock for the demand of intercity bus along the favoured CSY highway, while the shock is less noticeable along the less favoured one.

In addition to the CSY and CY highways that link Chengdu and Chongqing, we also examine the effect of the HSR line opening along a closely affected highway that leads to a major tourist destination. Leshan is a popular tourist destination about 120 km from Chengdu. Its main draw as a tourist attraction from locals and foreigners alike is that it is the home of the Leshan Giant Buddha, which is the largest stone-carved Buddha in the world. The 71 meter tall stone statue of the Buddha built in the 8th century during the Tang dynasty, and the scenic areas around it have been listed as a World Heritage Site by UNESCO since 1996.

A HSR line connecting Leshan with Chengdu was opened in Dec 2014 (see Table 1). An analysis of the effects of the opening of this HSR line provides us another opportunity to test for the HSR effect on pollution in a similar vicinity. We first use a navigation system to find out the optimal route between Chengdu to Leshan. This is demonstrated in Figure 1. Our formal regression results are presented in Panel B of Table 5, which are based on estimating the two-stage augmented RDIT model and using 26th December 2014 as the opening date.

The estimates suggest a reduction in air pollution for the area along the Chengdu-Leshan highway, similar to what was found for the CSY and CY highways. However, the magnitude of the coefficients is in general about half the size for this route. This could be the case for two possible reasons. First, using the Chengdu-Chongqing HSR saves more travelling time than using the Chengdu-Leshan HSR, and passenger load between Chengdu and Chongqing is much greater than for Chengdu to Leshan. Therefore, it is expected that there will be more passengers that switch from using road vehicles to HSR from Chengdu to Chongqing. Second, Leshan has various travel attractions and a large number of tourists. It would be more convenient to visit with a road vehicle. Many

tourists in Chengdu, for example, typically sign up for a one-day tour of the Leshan scenic area. This costs around 25 USD and includes a meal, transportation and tourist guide fees to view 4 scenic spots in the scenic area. This low-price deal is not feasible or profitable if the travel agency offers HSR as a travel mode from Chengdu to Leshan.

Temporal heterogeneity across affected highways — Highway use will vary by time of the day and day of the week. Traffic is expected to be more congested during peak hours. In this section, we exploit the availability of the high-frequency hourly data set and focus on examining intra-day variation in the effect of the HSR line opening on pollution. To do so, we first split our sample into 3-hour time periods (0-2 am, ... , 3-5 pm, ... , 9-11 pm). Next, we regress each time period for each line separately. Figure A6 plots the augmented RDiT estimates for the CSY highway. It shows that, as expected, the significant and highest reduction is observed in peak hours (3-5 pm) along the CSY highway. Lower levels of reduction are observed in morning and the effect is insignificant and approaches to zero during times of the day (after midnight to 5 am) when the HSR is not in operation. These results are suggestive of commuters substituting road vehicles for HSR during peak hour travel.

5.4 Estimates Based on DID

Table 6 presents the results of the DID analysis.²⁷ The results from the DID and RDiT approaches show some differences in the magnitudes of the estimated effects. This is not surprising as the two methods utilise different control group selection criteria and different identifying assumptions. However, it can be seen that the negative DID estimate for CO supports the conclusion previously presented that was based on using the augmented RDiT model, suggesting that the opening of HSR helped to reduce intercity air pollution.

²⁷We assess the common trends assumption and present the results for the event study analysis for air pollutants in Figure A5.

5.5 Robustness

This section conducts three different robustness checks involving spatial and temporal sensitivity tests to help ensure the validity of our estimates.²⁸

Donut-test and Bandwidth Selection — We perform two robustness checks regarding the cut-off day and our choice of bandwidth. The HSR line opening date was determined prior to the completion of the HSR line and therefore the cut-off day (or treatment date) is unlikely to be manipulated. To validate this fact, the first exercise is a donut test that drops observations around the cut-off day. For example, columns (7) - (9) of Table 3 show the RD estimated results that drop observations between 1, 7 and 21 days around the cut-off. The second exercise is to check whether our results are responsive to our choice of bandwidth. Figure A7 plots estimated coefficients and their corresponding 95% confidence intervals by repeatedly estimating augmented RDIT models using different bandwidths around the bandwidth that we use (365 days). This exercise confirms that our baseline estimates remain steady and is not sensitive to bandwidth choice.

Confounding Industrial Activities — A concern regarding the validity of baseline estimates is that a reduction in air pollution can be explained by reduction in industrial activities along affected intercity highways. To examine this possibility requires data for change in firm clusters over our sample period. We web scrape coordinates of industrial firms based on their names using Baidu Map API, using the latest annual survey of industrial manufacturing firms (ASIF) available to us which is from 2013. We next utilise a series of high spatial resolution VIIRS nightlight data to proxy for industrial activities along the highways. Appendix Figure A8 depicts a high correlation between nightlight intensity and firm clusters over space. Figure A9 shows that industrial production activities along the affected highways increased over the period 2015 to 2017, which will lead

²⁸The placebo exercise is unfortunately not feasible due to the limited temporal coverage of our air pollution data. Our empirical estimation uses pre-post two-year data relative to the treatment. Therefore, a proper placebo in our augmented time-disruptive RD design needs to randomly assign a false treatment at least two years prior to treatment. This would mean that the air pollution data need to cover periods prior to 2011 which is outside the boundary of the data set.

to a high level of air pollution. This would suggest that industrial activities are not a channel through which the HSR effect took place and further implies that our estimated coefficients are in the lower bound.

Excluding air pollution data within the Chongqing border — Chongqing is a large city and its urban centre (in the middle of the Chongqing cityloop) is 100 km away from its western border. An intercity passenger driving from Chengdu to Chongqing can stop by at any location between the Chongqing border and the Chongqing cityloop. In this section, we are interested in assessing whether the HSR reduction effects are mainly explained by highway segments within the Chongqing administrative border. Table A2 reports estimates by dropping air pollution grid cells within the Chongqing border for both the CSY and CY highways. As we continue to observe a reduction in air pollutants after the opening of HSR line for the highway road segments not within the Chongqing border, this suggests that the HSR line opening affected commuters travelling between Chengdu and Chongqing, and not just within Chongqing.

5.6 Cost-Benefit Analysis

The opening of a HSR line has many important economic implications for improving the welfare of residents and intercity passengers. Our cost-benefit analysis will mainly focus on two aspects: improved health benefits for residents who live along the main affected highways (due to reduced air pollution levels), and the value of time saved for commuters as a result of using the faster HSR as a mode of intercity travel.

Previous studies have documented health benefits from having lower levels of air pollution. To calculate precise estimates of the effects of the opening of the Chengdu-Chongqing HSR line for the nearest townships to the affected highways, the first step is to obtain information regarding the affected population in these areas. We take advantage of the availability of geo-coded China 2010 township population census data from the NASA Administrative Unit Centre Points with Population Estimates Programme (v4.11).

Figure 5 illustrates the affected townships along the highways and population information in the towns that are of interest. Altogether, the affected highways in our study pass through 120 towns in 25 counties. The CSY highway affects 1,503,244 residents and the CY highway affects 2,341,775 residents, as illustrated in Figure A10.

One important aspect of reduced air pollution levels is the health benefits gained from reduced exposure to particulate matter such as PM_{10} . To quantify the benefit gains from a reduction in exposure to these coarse total suspended particles, we use the estimates in Ebenstein et al. (2017)'s study, which shows that a $10 \mu\text{g}/\text{m}^3$ reduction in PM_{10} exposure would save 0.64 life years per Chinese resident. Using their estimates, we are able to monetise our estimated coefficients in Table 4 into dollar values. We found a reduction of 6.2% in PM_{10} for the CSY highway. The mean value of PM_{10} for the CSY highway prior to the HSR line opening is $79.83 \mu\text{g}/\text{m}^3$, so this corresponds to a $4.95 \mu\text{g}/\text{m}^3$ reduction in PM_{10} along the CSY highway. Therefore, the total life years gained for residents in the area along the CSY highway is estimated to be 476,228 years ($= (4.95/10) \times 0.64 \times 1,503,244$). Assuming we value each life year in China at 11,503 USD (Butt et al., 2019), the economic value of the life expectancy gain totals 5.48 billion USD.

Instead of the coarser PM_{10} , one can alternatively consider the effects of exposure to the finer $PM_{2.5}$ particles as exposure to it is also lethal and can lead to premature death. More estimates are available regarding the effects of changes in $PM_{2.5}$ on mortality in the literature, allowing benefit values to be estimated in terms of mortality rather than quality-adjusted life years (QALY). Our results show that the opening of the Chengdu-Chongqing HSR leads to a 9.8% reduction in $PM_{2.5}$ concentration after one year of operation. This is equivalent to a reduction of $4.72 \mu\text{g}/\text{m}^3$ based on the mean $PM_{2.5}$ concentration level one year prior to the opening date. A recent study by He et al. (2020) finds that a $10 \mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ leads to a 3.25% increase in the mortality rate. They also show that the average mortality rate in their treated provinces (including Chengdu and Chongqing) is 0.641% using 2015 mortality data from the China Centre for Disease

Control and Prevention. Applying these numbers in our context, this suggests that there is a 1.53% decrease in the mortality rate for residents who live along the CSY highway ($= 4.72/10 \times 3.25\%$). Using 9,636 ($=1,503,244 \times 0.641\%$) as the number of deaths from all causes in a year for residents who live along the highway, this implies that 147 premature deaths ($= 9,636 \times 1.53\%$) can be prevented. This is equivalent to an approximate social value of 169 million USD per year, using the average social value of statistical life (VSL) for a Chinese citizen to be 1.15 million USD in 2015 by [Fan et al. \(2020\)](#)'s calculation.²⁹

The HSR line opening can also have a beneficial effect on welfare in terms of the value of time saved by commuters. The Chengdu-Chongqing HSR line reduced intercity travelling time by 2.5 hours (see [Table 1](#)). Using an estimate that the average number of daily Chengdu-Chongqing HSR commuters is reported to be about 49,000 people each day, this therefore amounts to a total saving of 44,712,500 hours or 5,104 years per annum. Therefore, the time saved by marginal commuters can be monetised as an additional 58.71 million USD each year ($=5,104 \times 11,503$), in addition to the benefits of the improvements in air quality. Over time, these benefits can accrue quickly.

Our calculation is likely to represent a lower bound estimate of the full benefit of enhanced air quality as other health and non-health benefits are not accounted for, such as a reduction in CO or traffic accidents.^{30,31}

²⁹[Yang et al. \(2016\)](#) suggest similar estimates for VSL in China (1.13 million USD) calculated using survey data over the period of 2014 to 2015 for residents in Nanjing.

³⁰The effects of a reduction in CO in the literature have mainly focused on infant mortality and the results are relatively mixed. For example, [Currie et al. \(2009\)](#) found that a one ppm reduction in CO reduced 18 infant deaths per 100,000 births for New Jersey in the US. Similarly, [Knittel et al. \(2016\)](#) suggested an estimate of 82 avoided deaths per 100,000 live births for California over the period 2002-2007. In contrast, the estimated effect in Mexico was found to be much larger at 166 avoided deaths per 100,000 infant lives ([Arceo et al., 2016](#)). As China is somewhere in between Mexico and the US in terms of development status, we can use 92 infant deaths per 100,000 born to analyse the benefits for residents along the CSY highway. For these residents, our estimates suggest that an 8% reduction from the mean value of CO is equivalent to 0.07 mg/m³ reduction ($=0.853 \text{ mg/m}^3 \times 8\%$). Noting that for CO one ppm = 1.25 mg/m³ at standard pressure, we can calculate that this is equivalent to a reduction of 77 infant deaths ($=0.07/1.25 \times 92/100,000 \times 1,503,244$) during the first year of the operation of the Chengdu-Chongqing HSR line.

³¹An additional side benefit of reducing particulate matter is an improvement in road visibility, whose benefits are harder to quantify. Particulate matter during humid seasons can lead to the formation of foggy weather. Using thermal inversion as an instrument, [Sager \(2019\)](#) found that a 1 µg/m³ increase in PM_{2.5} per day in the UK could lead to an increase of 0.3 - 0.6% in the number of traffic accidents.

Overall, the economic benefits for CSY highway residents discussed in this section are in terms of the positive health benefits from having lower levels of particulate matter and from having faster commute times. The estimated benefits that have been calculated appear to be sizeable in proportion to the budgeted cost of 5.71 billion USD for building the Chengdu-Chongqing HSR line. Our calculation can be interpreted as suggestive evidence of the benefits of HSR that can provide. It can help social planners and policy makers in other countries make more informed decisions when they consider whether or not to make investments in green infrastructure projects such as HSR.

6 Conclusion

Green infrastructure such as HSR which automobile travellers can use as a substitute mode of transport can play a major role in reducing air pollution in the environment. However, HSR requires enormous investments and the net environmental effect resulting from a transition from motor vehicles to HSR for intercity travel is unclear. This paper investigates how the opening of a major HSR line that connects two of the biggest cities in China affects air pollution along the affected highway routes.

One empirical challenge is that it is difficult to know if changes in air quality over time are due to HSR opening or influenced by weather. This challenge is particularly serious when fitting a time-series regression discontinuity model. This is because estimation relies on observations that are far away from the treatment date in order to gain sufficient statistical power. In our case, the functional form of the interactions between weather and air pollution is hard to specify correctly. A failure to address this issue can result in misleading effects that are estimated. We overcome this issue of functional form by establishing counterfactual air pollutant concentration levels using a machine-learning algorithm that removes the confounding weather conditions. These predicted concentrations are later fed into an augmented regression discontinuity in time model to calculate

the local average treatment effects around both sides of the threshold. Our identification strategy involving machine-learning is not possible to implement without pollution data at the high-frequency and high spatial resolution level.

This paper provides two key findings. First, on average, the HSR line opening leads to a reduction in air pollution by 7.6%. This is most likely due to a substitution between automobile vehicle use and HSR. We also find that the largest pollution reduction effect is seen around the afternoon peak hour. Second, there are heterogeneous effects by pollutant. An examination of the results by pollutant shows that combustion air pollution such as those arising from petroleum emissions, CO, and diesel emissions, NO₂, decreased by 8% and 3% respectively along the affected highways. There was also a corresponding decline in the levels of particulate matter. We also find, however, an increase in ozone emissions. The increase in ozone emissions is likely due to the reduction in NO₂ levels relative to concentration levels one year prior to the HSR line opening.

Our results of a reduction in air pollution are strongest for the main highway between the two mega cities and smaller for alternative highways that involve relatively less passenger traffic. Our results are robust to various robustness tests and these findings are in alignment with the research on green infrastructure, which document that transit rail has led to a reduction of between 2-14% in air pollution.³²

Our results have two-fold implications. First, in recent years, China has invested trillions of dollars on transit infrastructure to improve road efficiency and boost economic development. Reduction in air pollution is an important benefit in terms of social welfare. Our results suggest that benefits to investments in transport infrastructure can likely be generalised to other heavily populated megacities suffering from deteriorating air quality. In this sense, HSR investment can be a useful tool to increase traffic efficiency, while at the same time allowing citizens to receive health benefits from improved air quality.

³²Guo et al. (2020)'s study suggests a reduction of 4.3% in average CO levels as a result of the opening of HSR lines. Our estimates are relatively higher than Guo et al. (2020)'s results due to a different setting but are still in alignment with the range of estimates in the green infrastructure HSR literature.

Our results here pertain to a context where there is high traffic volume along the highway leading toward big cities with clear city loops within which economic activities are concentrated, and where there can be a strong substitution effect for an alternative mode of travel that is affordable and quicker (see the city loops of Chengdu and Chongqing in Figure 1). However, we suggest caution in generalising our findings to other cities, such as cities with a more widely dispersed urban centre and larger peripheral areas as transit patterns might be quite different.

Second, in contrast with the existing literature (Chen and Whalley, 2012; Guo et al., 2020), our results suggest that the opening of a HSR line might not effectively bring down levels of ozone pollution over time. In fact, it might be even worse for areas along the affected highway, suggesting the need for complementary policies aimed at reducing O₃ pollution.³³ Whether our results are more generally applicable to more HSR lines in China, especially those that have lower ridership frequencies, requires further assessment.

Our work contributes to helping policy makers get a better grasp of the environmental benefits associated with HSR. While a greater reliance on rail has the potential to cut the growth of atmospheric pollutants, emissions from railway construction and maintenance must also be taken into account in determining whether HSR can help reduce emissions of gas pollutants over time.

³³See Wang et al. (2019) for more discussion on ozone pollution in China.

References

- Anderson, Michael L. (2014), "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion." *American Economic Review*, 104, 2763–2796.
- Arceo, Eva, Rema Hanna, and Paulina Oliva (2016), "Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from Mexico city." *The Economic Journal*, 126, 257–280.
- Butt, Thomas, Gordon G Liu, David D Kim, and Peter J Neumann (2019), "Taking stock of cost-effectiveness analysis of healthcare in China." *BMJ global health*, 4, e001418.
- Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014), "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica*, 82, 2295–2326.
- Carlsaw, David C and Karl Ropkins (2012), "Openairan R package for air quality data analysis." *Environmental Modelling & Software*, 27, 52–61.
- Chen, Shuai, Paulina Oliva, and Peng Zhang (2017), "The effect of air pollution on migration: Evidence from China (No. w24036)." *National Bureau of Economic Research*.
- Chen, Yihsu and Alexander Whalley (2012), "Green infrastructure: The effects of urban rail transit on air quality." *American Economic Journal: Economic Policy*, 4, 58–97.
- Chester, Mikhail and Arpad Horvath (2010), "Life-cycle assessment of high-speed rail: The case of California." *Environmental research letters*, 5, 014003.
- Cole, Matthew A, Robert JR Elliott, and Bowen Liu (2020), "The impact of the Wuhan Covid-19 lockdown on air pollution and health: A machine learning and augmented synthetic control approach." *Environmental and Resource Economics*, 76, 553–580.
- Currie, Janet, Matthew Neidell, and Johannes F Schmieder (2009), "Air pollution and infant health: Lessons from New Jersey." *Journal of Health Economics*, 28, 688–703.
- Davis, Lucas (2008), "The effect of driving restrictions on air quality in Mexico city." *Journal of Political Economy*, 116, 38–81.
- Deschenes, Olivier, Huixia Wang, Si Wang, and Peng Zhang (2020), "The effect of air pollution on body weight and obesity: Evidence from China." *Journal of Development Economics*, 102461.
- Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan (2013), "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India." *The Quarterly Journal of Economics*, 128, 1499–1545.
- Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng Zhou (2017), "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*, 114, 10384–10389.
- Fan, Maoyong, Guojun He, and Maigeng Zhou (2020), "The winter choke: coal-fired heating, air pollution, and mortality in China." *Journal of Health Economics*, 71, 102316.
- Gallego, Francisco, Juan-Pablo Montero, and Christian Salas (2013), "The effect of trans-

- port policies on car use: Evidence from Latin American cities." *Journal of Public Economics*, 107, 47–62.
- Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner (2021), "Subways and urban air pollution." *American Economic Journal: Applied Economics*, forthcoming.
- Givoni, Moshe and Frédéric Dobruszkes (2013), "A review of ex-post evidence for mode substitution and induced demand following the introduction of high-speed rail." *Transport Reviews*, 33, 720–742.
- Grange, Stuart K and David C Carslaw (2019), "Using meteorological normalisation to detect interventions in air quality time series." *Science of The Total Environment*, 653, 578–588.
- Grange, Stuart Kenneth, David Carlin Carslaw, Alastair Lewis, Eirini Boleti, and Christoph Heuglin (2018), "Random forest meteorological normalisation models for Swiss PM10 trend analysis." *Atmospheric Chemistry and Physics Discussions*.
- Gu, Yizhen, Chang Jiang, Junfu Zhang, and Ben Zou (2021), "Subways and road congestion." *American Economic Journal: Applied Economics*, 13, 83–115.
- Guo, Xiaoyang, Weizeng Sun, Shuyang Yao, and Siqi Zheng (2020), "Does high-speed railway reduce air pollution along highways? Evidence from China." *Transportation Research Part D: Transport and Environment*, 89, 102607.
- Hausman, Catherine and David S Rapson (2018), "Regression discontinuity in time: Considerations for empirical applications." *Annual Review of Resource Economics*, 10, 533–552.
- He, Guojun, Tong Liu, and Maigeng Zhou (2020), "Straw burning, PM2.5, and death: Evidence from china." *Journal of Development Economics*, 145, 102468.
- Imbens, Guido and Karthik Kalyanaraman (2012), "Optimal bandwidth choice for the regression discontinuity estimator." *The Review of Economic Studies*, 79, 933–959.
- Knittel, Christopher R, Douglas L Miller, and Nicholas J Sanders (2016), "Caution, drivers! children present: Traffic, pollution, and infant health." *Review of Economics and Statistics*, 98, 350–366.
- Kong, Lei, Xiao Tang, Jiang Zhu, Zifa Wang, Jianjun Li, Huangjian Wu, Qizhong Wu, Huansheng Chen, Lili Zhu, Wei Wang, et al. (2021), "A 6-year-long (2013–2018) high-resolution air quality reanalysis dataset in china based on the assimilation of surface observations from cnemc." *Earth System Science Data*, 13, 529–570.
- Lalive, Rafael, Simon Luechinger, and Armin Schmutzler (2018), "Does expanding regional train service reduce air pollution?" *Journal of Environmental Economics and Management*, 92, 744–764.
- Lang, Jianlei, Shuiyuan Cheng, Ying Zhou, Yonglin Zhang, and Gang Wang (2014), "Air pollutant emissions from on-road vehicles in China, 1999–2011." *Science of the Total Environment*, 496, 1–10.
- Li, Pei, Yi Lu, and Jin Wang (2020), "The effects of fuel standards on air pollution: Evidence from China." *Journal of Development Economics*, 102488.

- Li, Shanjun, Yanyan Liu, Avralt-Od Purevjav, and Lin Yang (2019), "Does subway expansion improve air quality?" *Journal of Environmental Economics and Management*, 96, 213–235.
- Ren, Xiaohong, Fang Wang, Chunyang Wang, Zongyang Du, Zhenhua Chen, Jiamei Wang, and Ting Dan (2019), "Impact of high-speed rail on intercity travel behavior change." *Journal of Transport and Land Use*, 12, 265–285.
- Sager, Lutz (2019), "Estimating the effect of air pollution on road safety using atmospheric temperature inversions." *Journal of Environmental Economics and Management*, 98, 102250.
- Shi, Zongbo, Congbo Song, Bowen Liu, Gongda Lu, Jingsha Xu, Tuan Van Vu, Robert JR Elliott, Weijun Li, William J Bloss, and Roy M Harrison (2021), "Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns." *Science Advances*, 7, eabd6696.
- Vu, Van Tuan, Zongbo Shi, Jing Cheng, Qiang Zhang, Kebin He, Shuxiao Wang, and Roy Harrison (2019), "Assessing the impact of clean air action on air quality trends in Beijing using a machine learning." *Atmos. Chem. Phys*, 19, 11303–11314.
- Wang, Chunchao, Qianqian Lin, and Yun Qiu (2020), "Productivity loss amid invisible pollution." Available at SSRN 3695499.
- Wang, Kun, Wenyi Xia, and Anming Zhang (2017), "Should china further expand its high-speed rail network? Consider the low-cost carrier factor." *Transportation Research Part A: Policy and Practice*, 100, 105–120.
- Wang, Nan, Xiaopu Lyu, Xuejiao Deng, Xin Huang, Fei Jiang, and Aijun Ding (2019), "Aggravating O₃ pollution due to NO_x emission control in eastern China." *Science of the Total Environment*, 677, 732–744.
- Wee, Bert van, Pieter Janse, and Robert van Den Brink (2005), "Comparing energy use and environmental performance of land transport modes." *Transport Reviews*, 25, 3–24.
- Yang, Zhao, Pan Liu, and Xin Xu (2016), "Estimation of social value of statistical life using willingness-to-pay method in Nanjing, China." *Accident Analysis & Prevention*, 95, 308–316.
- Yu, Bo, Wang-Sheng Lee, and Shuddhasattwa Rafiq (2020), "Air pollution quotas and the dynamics of internal skilled migration in Chinese cities (No. DP13479)." *IZA Discussion Paper*.
- Zhang, Junfeng Jim, Yongjie Wei, and Zhangfu Fang (2019), "Ozone pollution: A major health hazard worldwide." *Frontiers in Immunology*, 10, 2518–2518.
- Zheng, Siqi, Xiaonan Zhang, Weizeng Sun, and Jianghao Wang (2019), "The effect of a new subway line on local air quality: A case study in Changsha." *Transportation Research Part D: Transport and Environment*, 68, 26–38.

Table 1: High-Speed Railway Openings and Affected Highways

| <i>Panel A: Highways</i> | | | | |
|--------------------------|----------------|-------------------------|--------------|------------|
| Highway Name | Highway Length | Average Travelling Time | Opening Date | Toll (USD) |
| Cheng-Sui-Yu | 322 KM | 4hrs | Dec/2007 | \$20 |
| Cheng-Yu | 344 KM | 4hrs 30mins | Aug/1995 | \$20 |
| Chengdu-Leshan | 140 KM | 2hrs 10mins | Dec/1999 | \$7 |

| <i>Panel B: HSR</i> | | | | |
|-------------------------|------------|-------------------------|--------------|--------------------|
| High-Speed Railway Line | HSR Length | Average Travelling Time | Opening Date | Ticket Price (USD) |
| Chengdu-Chongqing HSR | 308 KM | 1hr 30mins | 26/Dec/2015 | \$17 - \$21 |
| Chengdu-Leshan HSR | 135 KM | 52 mins | 20/Dec/2014 | \$8 |

Table 2: Descriptive Statistics for the Cheng-Sui-Yu Highway pre- and post-Opening of the Chengdu-Chongqing HSR line

| VARIABLES | (1) (2) (3) (4) (5) One Year Before HSR opening | | | | | (6) (7) (8) (9) (10) One Year After HSR opening | | | | |
|---|--|----------|----------|---------|---------|--|----------|----------|----------|---------|
| | N | mean | sd | min | max | N | mean | sd | min | max |
| CO | 193,248 | 0.849 | 0.246 | 0.201 | 2.443 | 192,720 | 0.819 | 0.283 | 0.174 | 2.697 |
| De-weathered CO | 193,248 | 0.853 | 0.164 | 0.498 | 1.535 | 192,720 | 0.822 | 0.183 | 0.471 | 1.457 |
| NO ₂ | 193,248 | 21.71 | 12.94 | 0 | 197.9 | 192,720 | 21.15 | 12.38 | 0 | 160.6 |
| De-weathered NO ₂ | 193,248 | 22.13 | 7.524 | 4.988 | 52.03 | 192,720 | 22.11 | 7.608 | 4.827 | 52.53 |
| PM _{2.5} | 193,248 | 46.93 | 29.36 | 0 | 295.5 | 192,720 | 45.31 | 29.47 | 0 | 411.9 |
| De-weathered PM _{2.5} | 193,248 | 48.11 | 19.22 | 24.48 | 119.8 | 192,720 | 45.86 | 16.68 | 17.93 | 123.3 |
| PM ₁₀ | 193,248 | 77.88 | 44.98 | 0 | 421.9 | 192,720 | 75.11 | 42.70 | 0 | 454.9 |
| De-weathered PM ₁₀ | 193,248 | 79.83 | 28.04 | 41.70 | 177.0 | 192,720 | 76.48 | 23.95 | 36.91 | 162.8 |
| O ₃ | 193,248 | 52.31 | 33.58 | 0 | 260.1 | 192,720 | 55.78 | 33.69 | 0 | 240.4 |
| De-weathered O ₃ | 193,248 | 53.05 | 26.19 | 9.537 | 138.9 | 192,720 | 55.58 | 26.81 | 11.79 | 141.1 |
| Ground temperature (Kelvin) | 193,248 | 291.7 | 7.718 | 271.5 | 309.1 | 192,720 | 291.5 | 8.412 | 268.6 | 310.2 |
| Relative humidity | 193,248 | 64.37 | 17.98 | 12.95 | 100 | 192,720 | 66.11 | 18.18 | 13.96 | 100 |
| Surface pressure (Pa) | 193,248 | 97,130 | 1,153 | 92,594 | 100,446 | 192,720 | 97,139 | 1,194 | 92,826 | 101,534 |
| Wind speed (m/s) | 193,248 | 3.961 | 2.237 | 0.0118 | 20.72 | 192,720 | 3.878 | 2.281 | 0.0111 | 17.24 |
| Wind direction (degree) | 193,248 | 200.3 | 68.52 | 0.00836 | 360.0 | 192,720 | 198.4 | 70.87 | 0.000338 | 360.0 |
| Precipitation (kg m ⁻² s ⁻¹) | 193,248 | 3.19e-05 | 0.000115 | 0 | 0.00266 | 192,720 | 2.94e-05 | 9.50e-05 | 0 | 0.00253 |
| Thermal inversion (Celsius) | 193,248 | 0.153 | 0.469 | 0 | 4.276 | 192,720 | 0.163 | 0.477 | 0 | 4.737 |

Note: All concentrations are in $\mu\text{g}/\text{m}^3$ except for CO which are in units of mg/m^3 .

Table 3: Augmented RDiT for Stacked Air Pollution along the Cheng-Sui-Yu Highway

| VARIABLES | Log Residual (Hourly Stacked Pollutant Concentration) | | | | | | Donut Test | | |
|--------------------|---|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| OpeningHSR | -0.110*** (0.007) | -0.110*** (0.007) | -0.128*** (0.007) | -0.059*** (0.004) | -0.058*** (0.004) | -0.076*** (0.004) | -0.074*** (0.004) | -0.082*** (0.004) | -0.092*** (0.004) |
| Day | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| OpeningHSR×Day | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| Lag Outcome Var. | 0.272*** (0.002) | 0.272*** (0.002) | 0.260*** (0.002) | 0.130*** (0.001) | 0.129*** (0.001) | 0.128*** (0.001) | 0.127*** (0.001) | 0.127*** (0.001) | 0.129*** (0.001) |
| Observations | 87,720 | 87,720 | 87,720 | 87,720 | 87,720 | 87,720 | 87,360 | 85,920 | 82,560 |
| R-squared | 0.263 | 0.262 | 0.251 | 0.125 | 0.125 | 0.125 | 0.124 | 0.125 | 0.130 |
| Data | Raw | Raw | Raw | Deweather | Deweather | Deweather | Deweather | Deweather | Deweather |
| Holiday FE 1st | NO | YES | YES | NO | YES | YES | YES | YES | YES |
| Day of Year FE 1st | NO | NO | YES | NO | NO | YES | YES | YES | YES |
| Window (days) | 365 | 365 | 365 | 365 | 365 | 365 | Drop 1 Day | Drop 7 Day | Drop 21 Day |

Note: Regression estimates are obtained from the second stage of the augmented RDiT approach. *OpeningHSR* is the variable of interest. All AQR grid cell observations are collapsed by highway line. Residuals in the second stage are obtained by regressing air pollutants on hour, day of year and pollutant FEs over a four-year window (2014 Jan - 2017 Dec) in the first stage. In the second stage, we estimate a local linear regression by regressing residuals on lagged air pollutants, the HSR opening date and time trend as a running variable on a narrower two-year window (2015 Jan - 2016 Dec). Regressions using raw data include weather controls in the first stage. 'Raw' indicates non machine-learning processed air pollutant data; 'Deweather' indicates machine-learning processed data. S.E. in parentheses is bootstrapped 500 times. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Augmented RDiT by Type of Pollutants for the Cheng-Sui-Yu Highway

| VARIABLES | Log Residual (Hourly Pollutant Concentration) | | | | | | | | | |
|------------------|---|--------------------------|------------------------|-----------------------|-------------------------|----------------------|--------------------------|------------------------|-----------------------|--------------------------|
| | (1) CO | (2) PM _{2.5} | (3) NO ₂ | (4) O ₃ | (5) PM ₁₀ | (6) CO | (7) PM _{2.5} | (8) NO ₂ | (9) O ₃ | (10) PM ₁₀ |
| OpeningHSR | -0.141*** (0.006) | -0.295*** (0.011) | -0.112*** (0.011) | -0.012 (0.008) | -0.266*** (0.010) | -0.080*** (0.002) | -0.098*** (0.004) | -0.030*** (0.005) | 0.012*** (0.002) | -0.062*** (0.003) |
| Day | 0.000*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| OpeningHSR × Day | -0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.001*** (0.000) | -0.000 (0.000) | 0.000*** (0.000) | 0.001*** (0.000) |
| Lag Outcome Var. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 |
| R-squared | 0.265 | 0.380 | 0.394 | 0.280 | 0.366 | 0.169 | 0.146 | 0.123 | 0.205 | 0.166 |
| Data | Raw | Raw | Raw | Raw | Raw | Deweather | Deweather | Deweather | Deweather | Deweather |
| Window (days) | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 |

Note: Regression estimates are obtained from the second stage of the augmented RDiT approach. *OpeningHSR* is the variable of interest. All AQR grid cell observations are collapsed by highway line. Residuals in the second stage are obtained by regressing air pollutants on hour FEs and day of Julian year FEs over a four-year window (2014 Jan - 2017 Dec) in the first stage. In the second stage, we estimate a local linear regression by regressing residuals on lagged air pollutants, the HSR opening date and time trend as a running variable on a narrower two-year window (2015 Jan - 2016 Dec). Regressions using raw data include weather controls in the first stage. 'Raw' indicates non machine-learning processed air pollutant data; 'Deweather' indicates machine-learning processed data. S.E. in parentheses is bootstrapped 500 times. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Augmented RDiT by Type of Pollutant for the Cheng-Yu and Chengdu-Leshan Highway

| VARIABLES | Log Residual (Hourly Pollutant Concentration) | | | | | | | | | |
|--|---|--------------------------|------------------------|-----------------------|-------------------------|----------------------|--------------------------|------------------------|-----------------------|--------------------------|
| | (1) CO | (2) PM _{2.5} | (3) NO ₂ | (4) O ₃ | (5) PM ₁₀ | (6) CO | (7) PM _{2.5} | (8) NO ₂ | (9) O ₃ | (10) PM ₁₀ |
| <i>Panel A: Cheng-Yu Highway</i> | | | | | | | | | | |
| OpeningHSR | -0.128*** (0.006) | -0.210*** (0.011) | -0.079*** (0.009) | 0.029*** (0.009) | -0.196*** (0.010) | -0.077*** (0.002) | -0.048*** (0.004) | 0.005 (0.004) | 0.032*** (0.002) | -0.051*** (0.004) |
| Day | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.000 (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| OpeningHSR × Day | -0.000*** (0.000) | 0.001*** (0.000) | -0.000 (0.000) | 0.001*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | 0.001*** (0.000) | 0.000* (0.000) | 0.000*** (0.000) | 0.001*** (0.000) |
| <i>Panel B: Chengdu-Leshan Highway</i> | | | | | | | | | | |
| OpeningHSR | -0.195*** (0.007) | -0.328*** (0.010) | -0.069*** (0.011) | 0.042*** (0.012) | -0.278*** (0.010) | -0.065*** (0.002) | -0.054*** (0.003) | -0.038*** (0.005) | 0.003 (0.003) | -0.031*** (0.003) |
| Day | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | -0.001*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | -0.001*** (0.000) | -0.000 (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| OpeningHSR × Day | -0.000 (0.000) | 0.000*** (0.000) | 0.000** (0.000) | 0.001*** (0.000) | 0.000* (0.000) | 0.000*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| Observations | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 |
| Lag Outcome Var. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Data | Raw | Raw | Raw | Raw | Raw | Deweather | Deweather | Deweather | Deweather | Deweather |
| Window (days) | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 |

Note: Regression estimates are obtained from the second stage of the augmented RDiT approach. *OpeningHSR* is the variable of interest. All AQR grid cell observations are collapsed by highway line. Residuals in the second stage are obtained by regressing air pollutants on hour FEs and day of Julian year FEs over a four-year window in the first stage (2014 Jan - 2017 Dec for the Cheng-Yu highway and 2013 Jan - 2016 Dec for the Chengdu-Leshan highway). In the second stage, we estimate a local linear regression by regressing residuals on lagged air pollutants, the HSR opening date and time trend as a running variable on a narrower two-year window (2015 Jan - 2016 Dec for Cheng-Yu highway and 2014 Jan - 2015 Dec for Chengdu-Leshan highway). Regressions using raw data include weather controls in the first stage. 'Raw' indicates non machine-learning processed air pollutant data; 'Deweather' indicates machine-learning processed data. S.E. in parentheses is bootstrapped 500 times. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Difference-in-Difference Estimates for the Cheng-Sui-Yu Highway

| VARIABLES | Log (deweathered pollutant concentrations) | | | | |
|---------------------------------|--|--------------------------|-------------------------|------------------------|-----------------------|
| | (1) CO | (2) PM _{2.5} | (3) PM ₁₀ | (4) NO ₂ | (5) O ₃ |
| Treated AQR \times OpeningHSR | -0.035*** (0.008) | -0.051** (0.015) | -0.004 (0.015) | 0.003 (0.011) | 0.014 (0.024) |
| AQR Grid Cell FE | Yes | Yes | Yes | Yes | Yes |
| Day-by-Year FE | Yes | Yes | Yes | Yes | Yes |
| Weather | Yes | Yes | Yes | Yes | Yes |
| Observations | 87,660 | 87,660 | 87,660 | 87,660 | 87,660 |
| R-squared | 0.965 | 0.971 | 0.977 | 0.968 | 0.970 |

Note: This table shows the DID estimates using deweathered daily pollution data (in logarithms) for the treated Cheng-Sui-Yu (CSY) highway. The opening of the Chengdu-Chongqing HSR line is the treatment and data are analysed over the period of 2014 Jan - 2017 Dec. Treated observations are AQR grid cells along the CSY highway. The control cells are observations within random road segments in Sichuan and Chongqing provinces and are road segments that drivers cannot directly use to commute between Chengdu and Chongqing. A third-order polynomial in weather controls are included to improve balancing between treated and control AQR grid cells. All regressions further control for AQR grid cells and interactive day by year fixed effects. Robust standard errors reported in parentheses are clustered by road segments. *** p<0.01, ** p<0.05, * p<0.1.

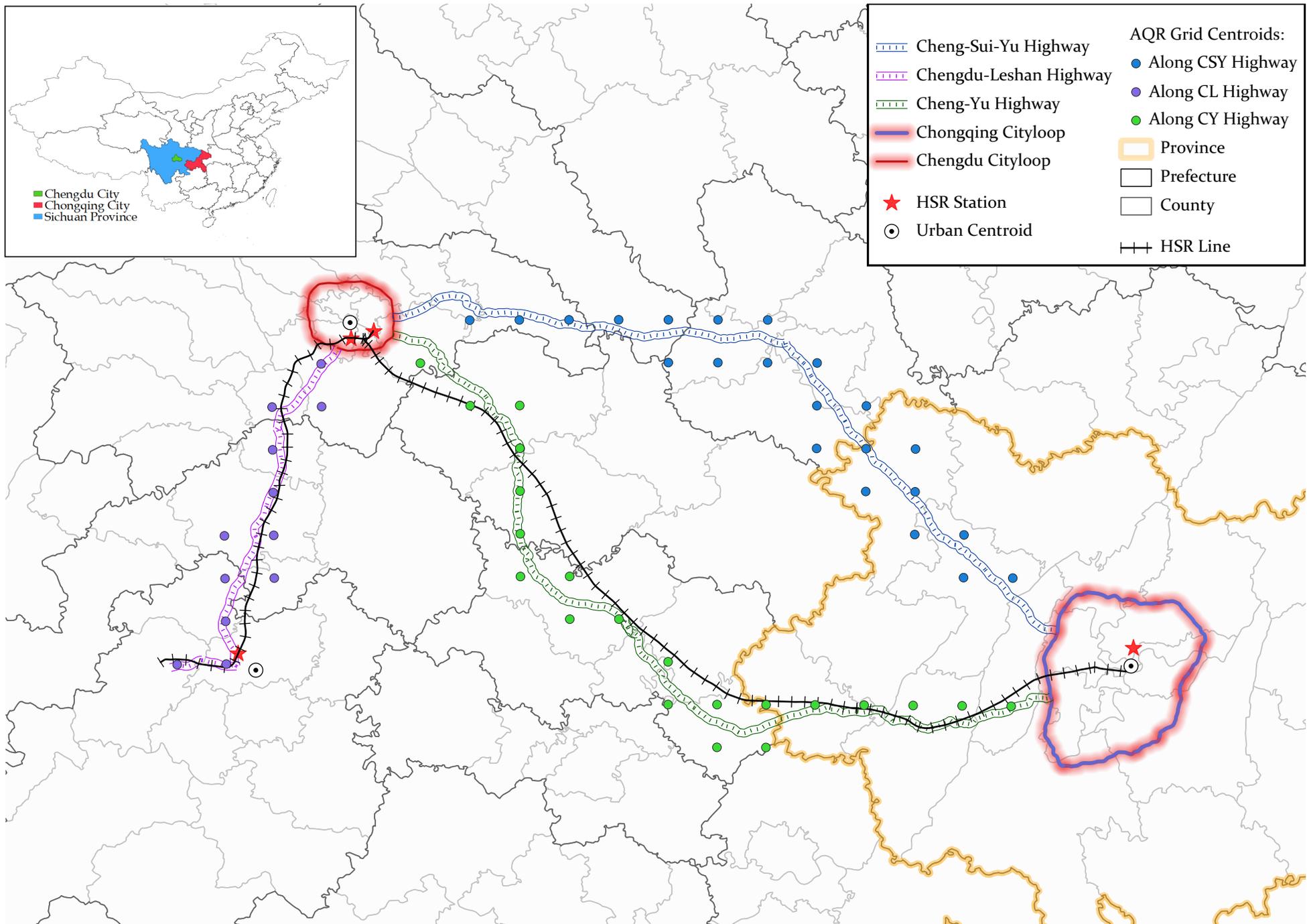


Figure 1: Illustration of the Air Pollution Grid Cells along the Chengdu-Chongqing HSR Line

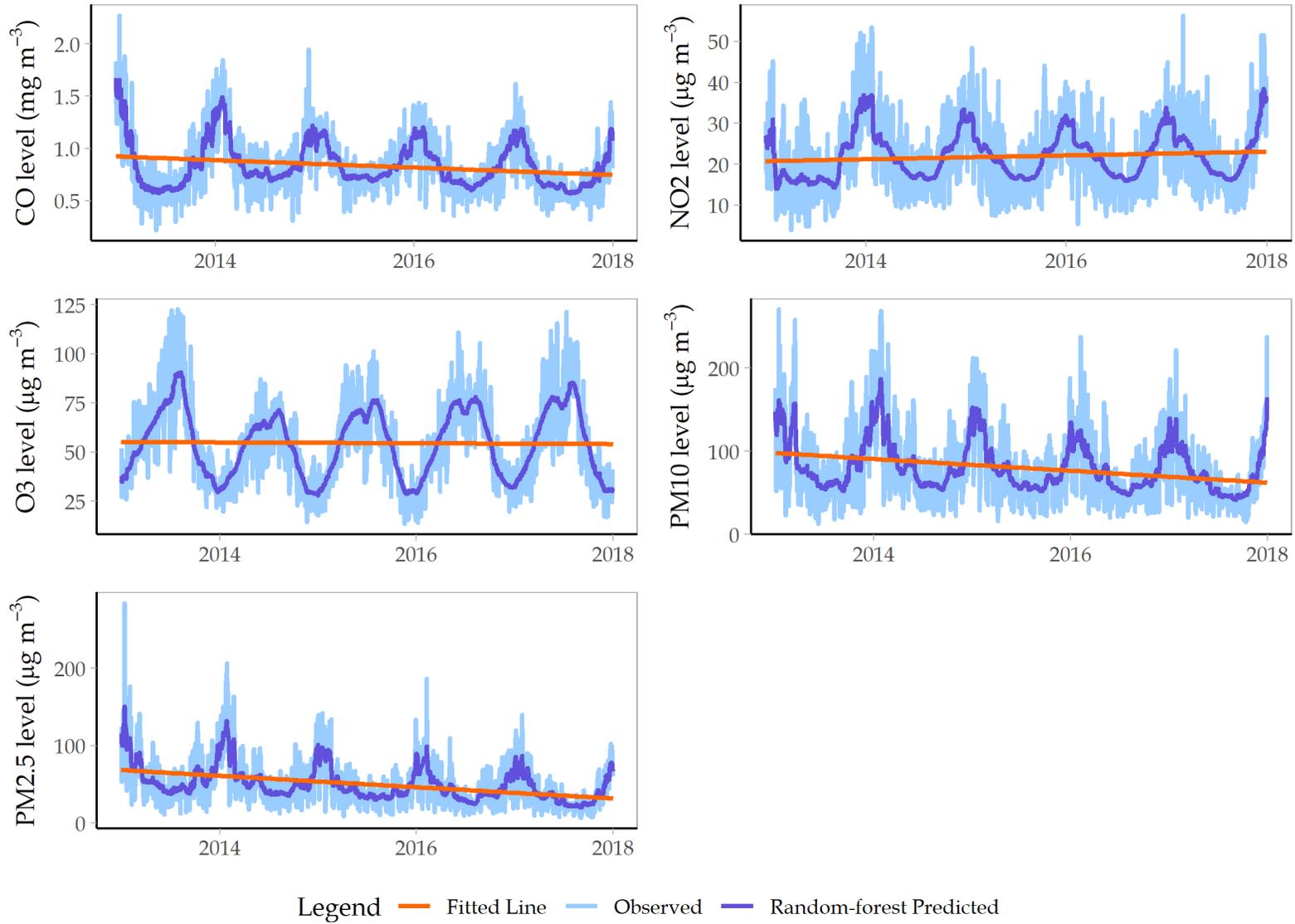


Figure 2: An Illustration of the Deweathering Process for the Cheng-Sui-Yu Highway

Note: This plot compares observed trends of hourly AQR data (collapsed by day) against trends after a random-forest process that removes weather effects for five pollutants along the Cheng-Sui-Yu highway.

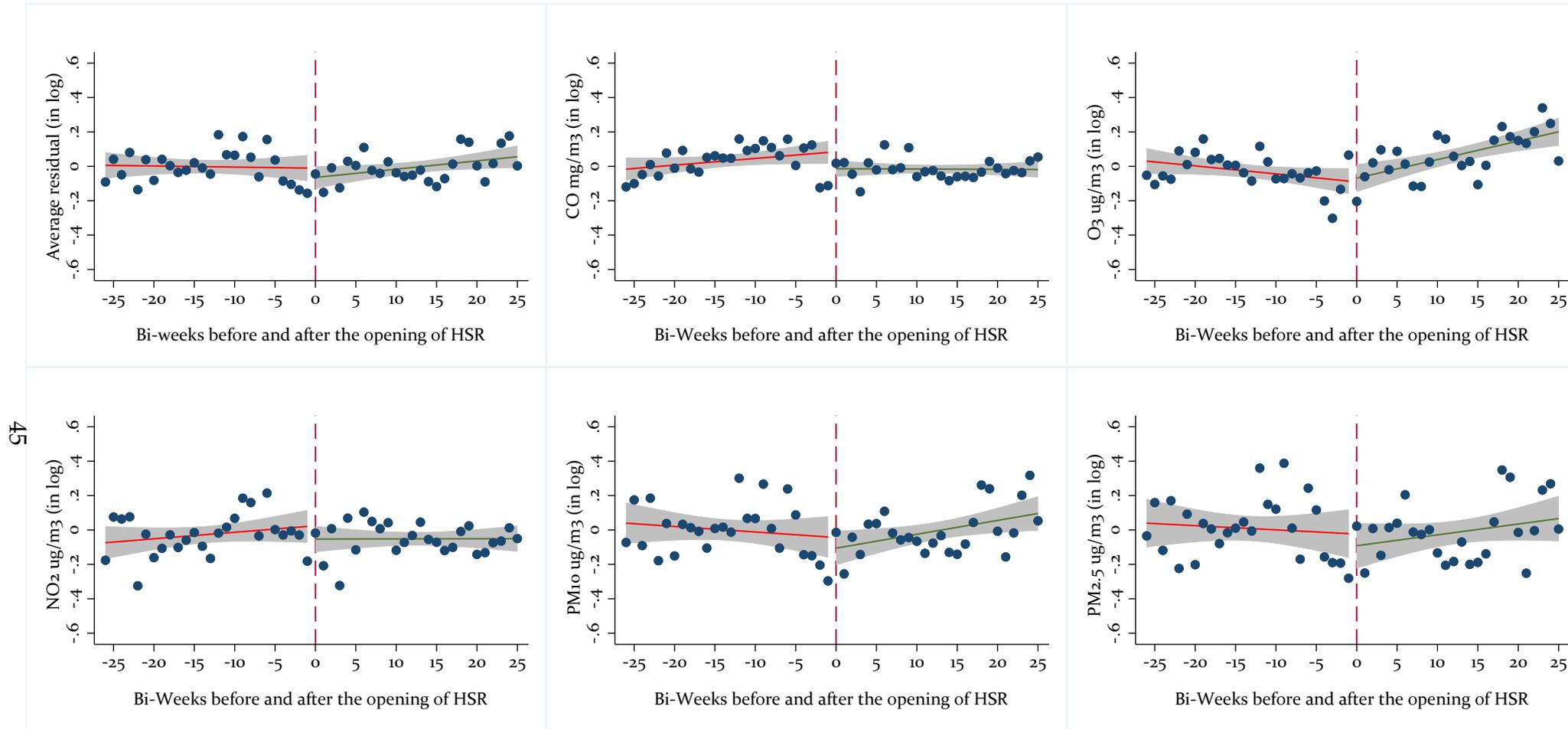


Figure 3: Effects of HSR on Air Pollution Concentration (by Type of Pollutant) for the Cheng-Sui-Yu Highway (Raw Data)

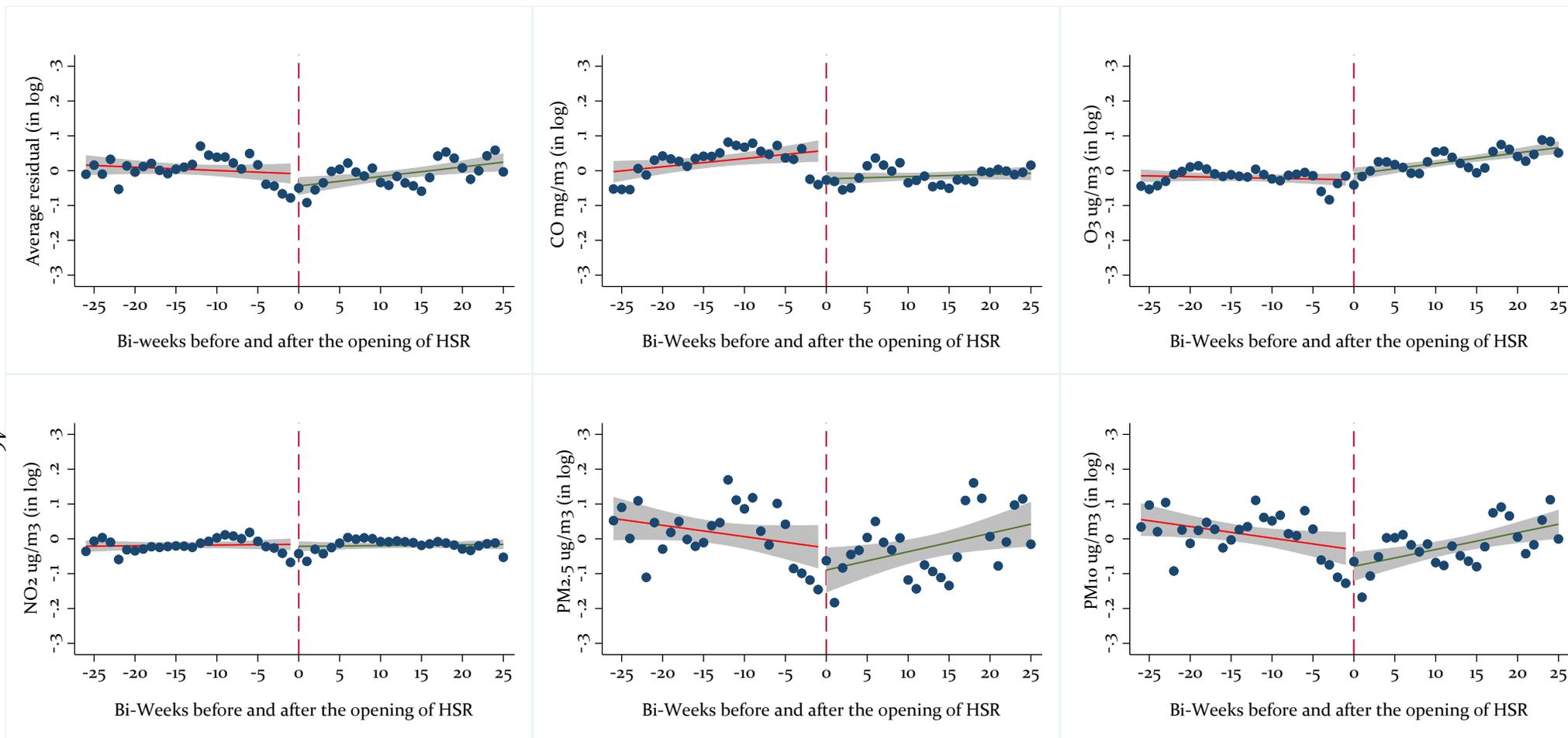


Figure 4: Effects of HSR on Air Pollution Concentration (by Type of Pollutant) for the Cheng-Sui-Yu Highway (Deweathered Data)

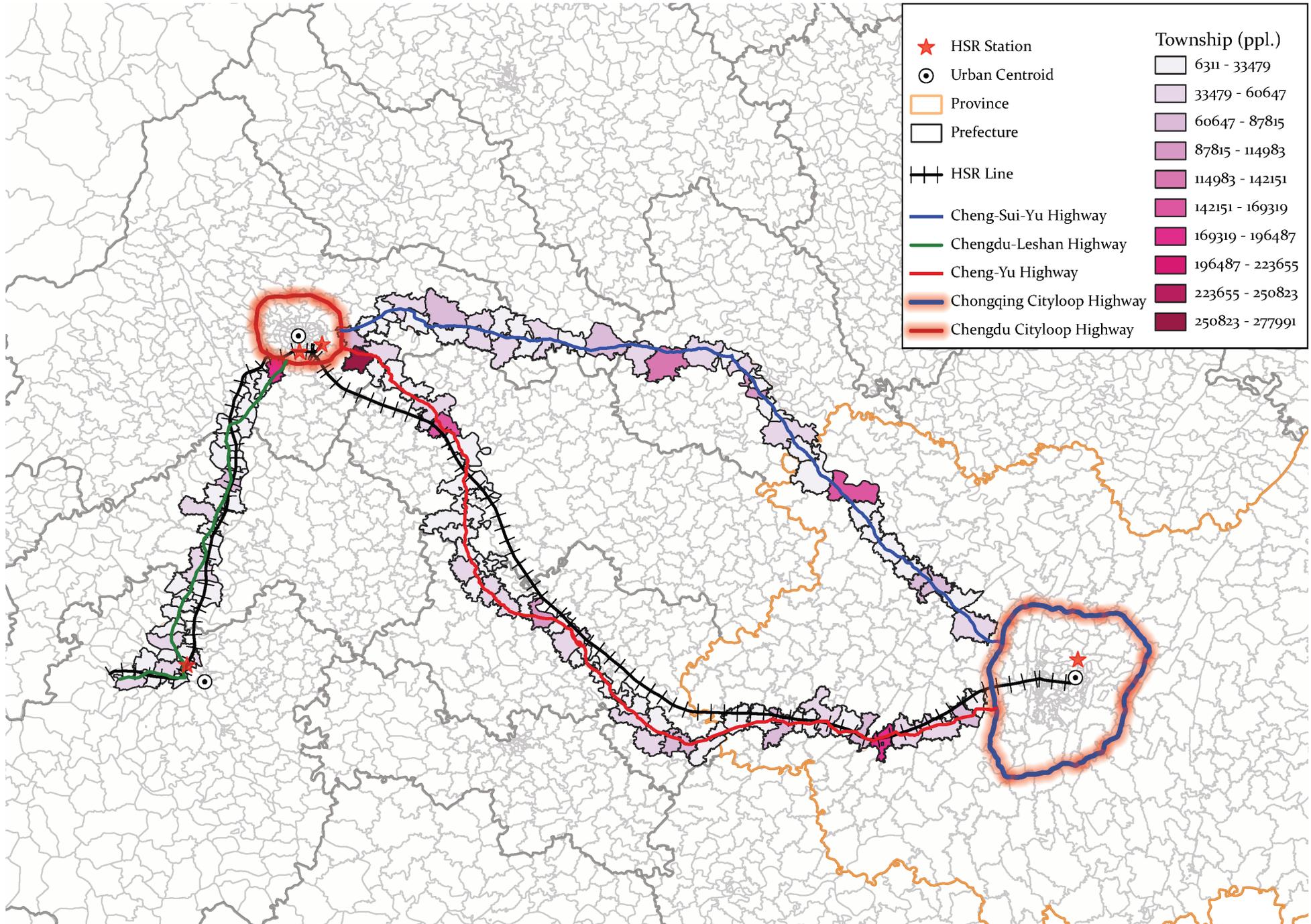


Figure 5: Township Population along the Affected Highways

Online Appendix (Not for Publication)

This page is intentionally left blank.

Additional Figures and Tables

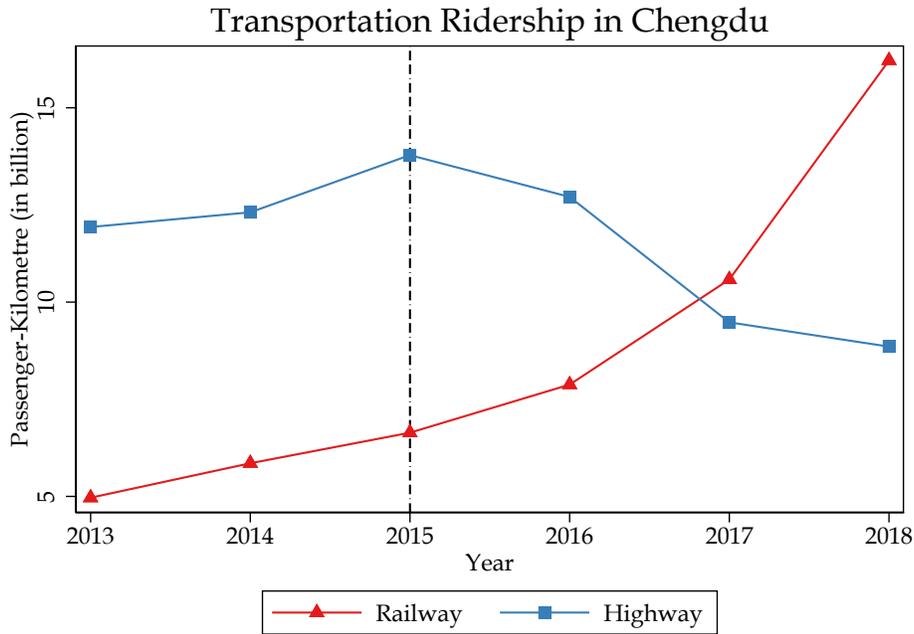


Figure A1: Time Trends of Highway Use and Railway Ridership in Chengdu (Source: *Sichuan Statistical Yearbooks* in Various Years)

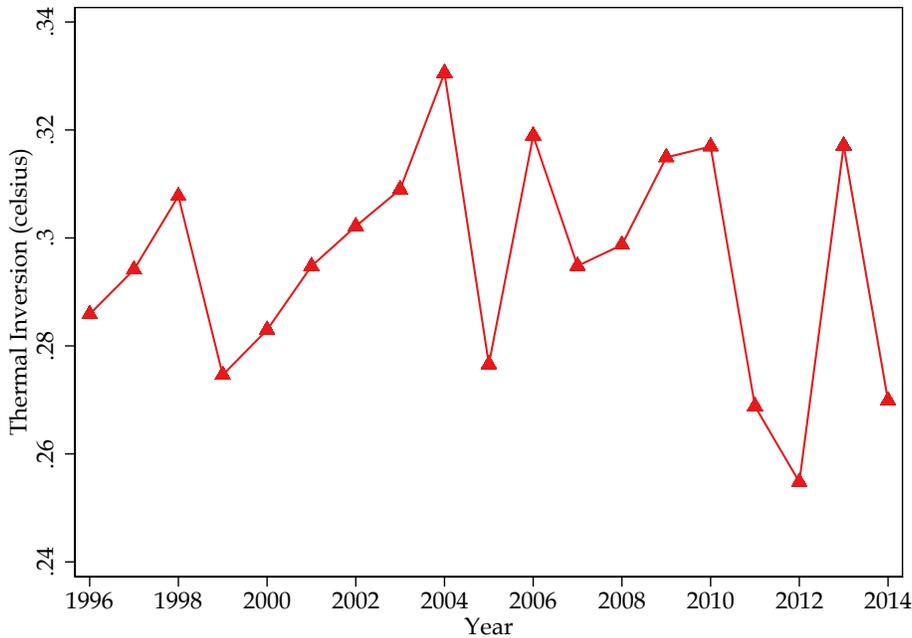


Figure A2: Thermal Inversion Trends

Note: This plot replicates thermal inversion in [Chen et al. \(2017\)](#) by collapsing 6 hourly NASA MERRA-2 data to yearly average across 2859 counties, downloaded from the Ministry of Civil Affairs of the P.R.C.. The data for the plot are obtained by searching for the altitude of each grid and calculating the temperature difference between two closest atmospheric layers from the nearest surface. We only keep positive values of thermal inversion and truncate negatives to 0.

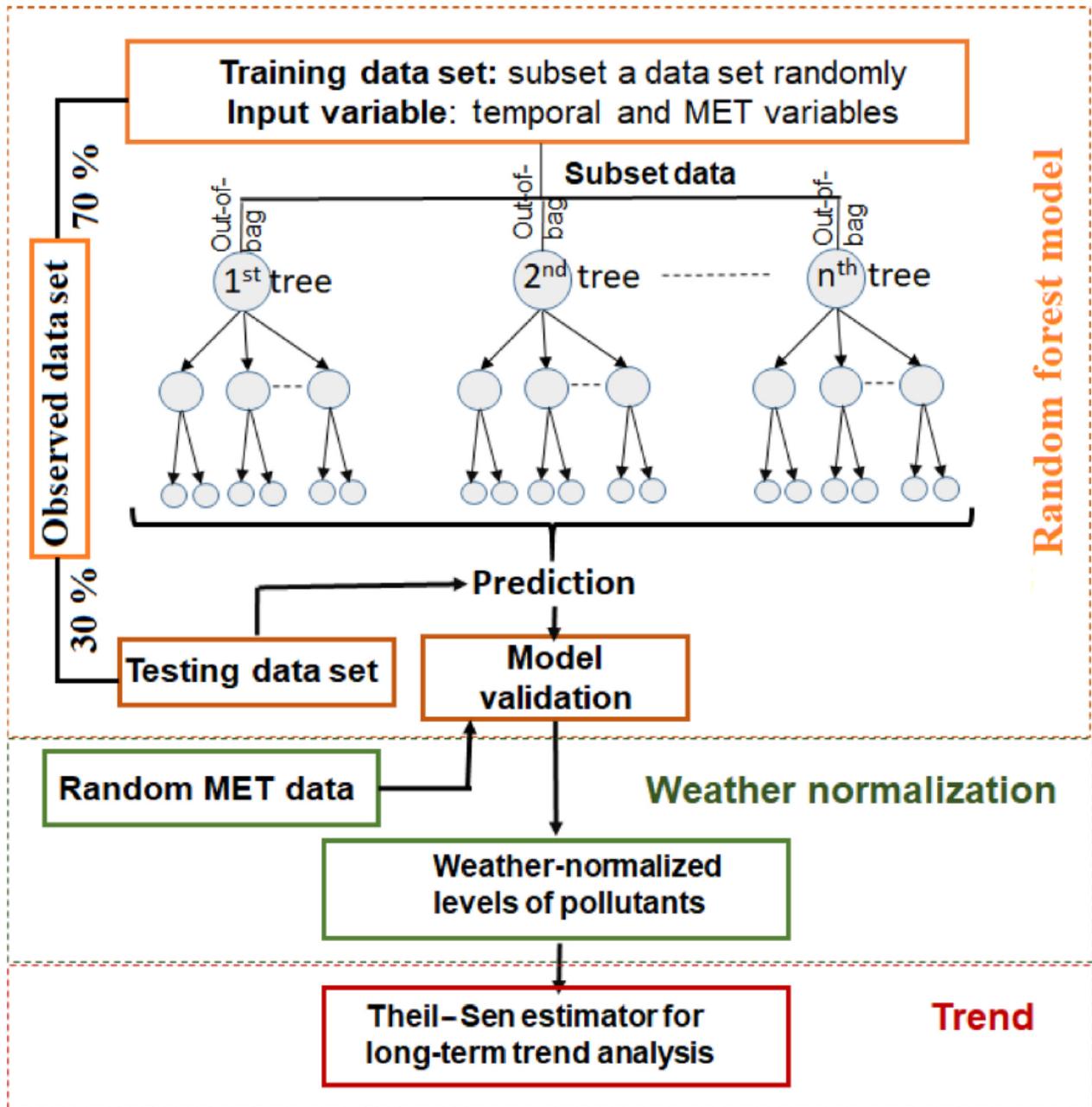


Figure A3: A Random-Forest Prediction Process for Deweathering Data (Source: [Vu et al., 2019](#))

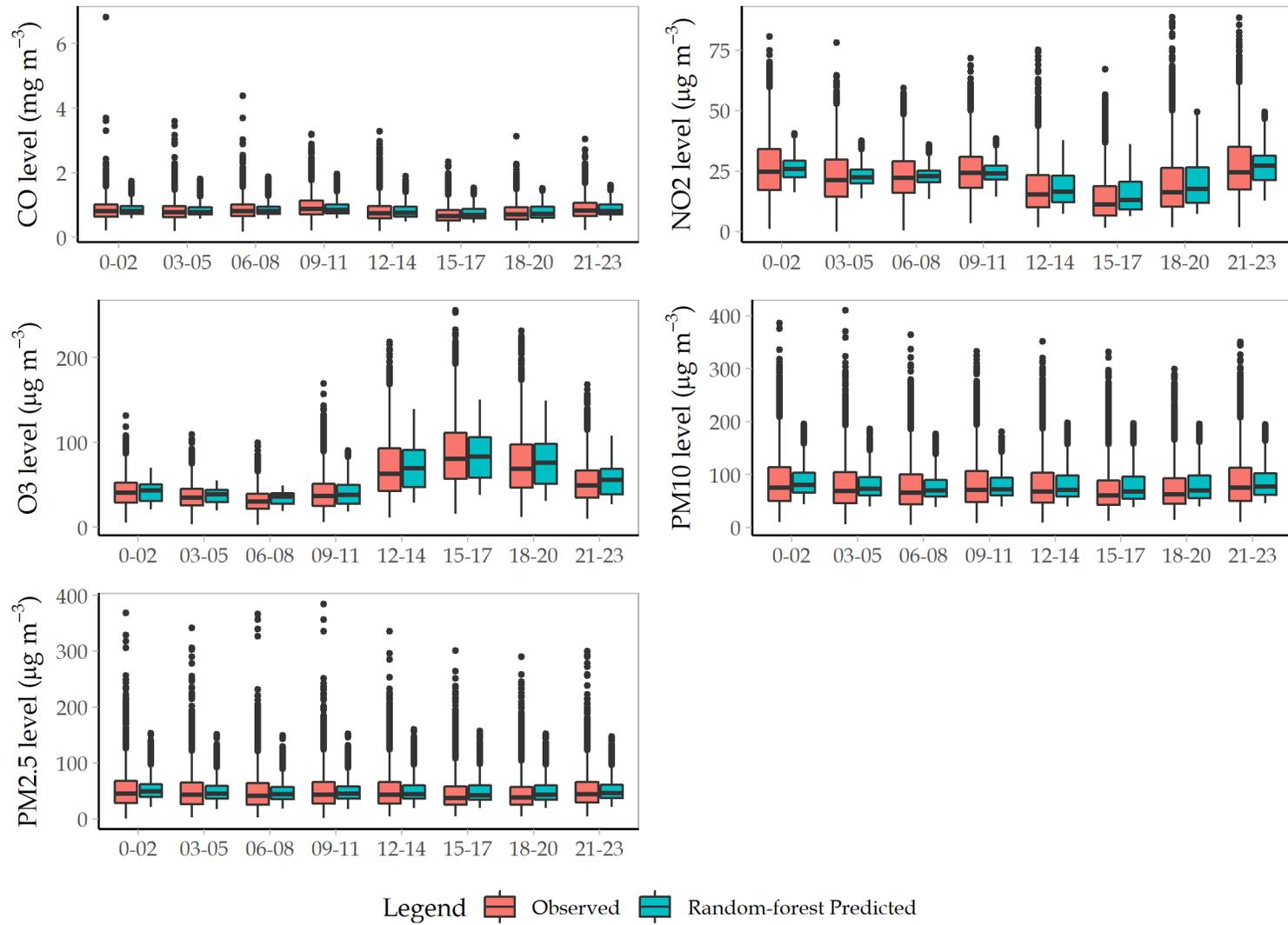


Figure A4: Box Plots for the Deweathering Process at Different Times in a Day

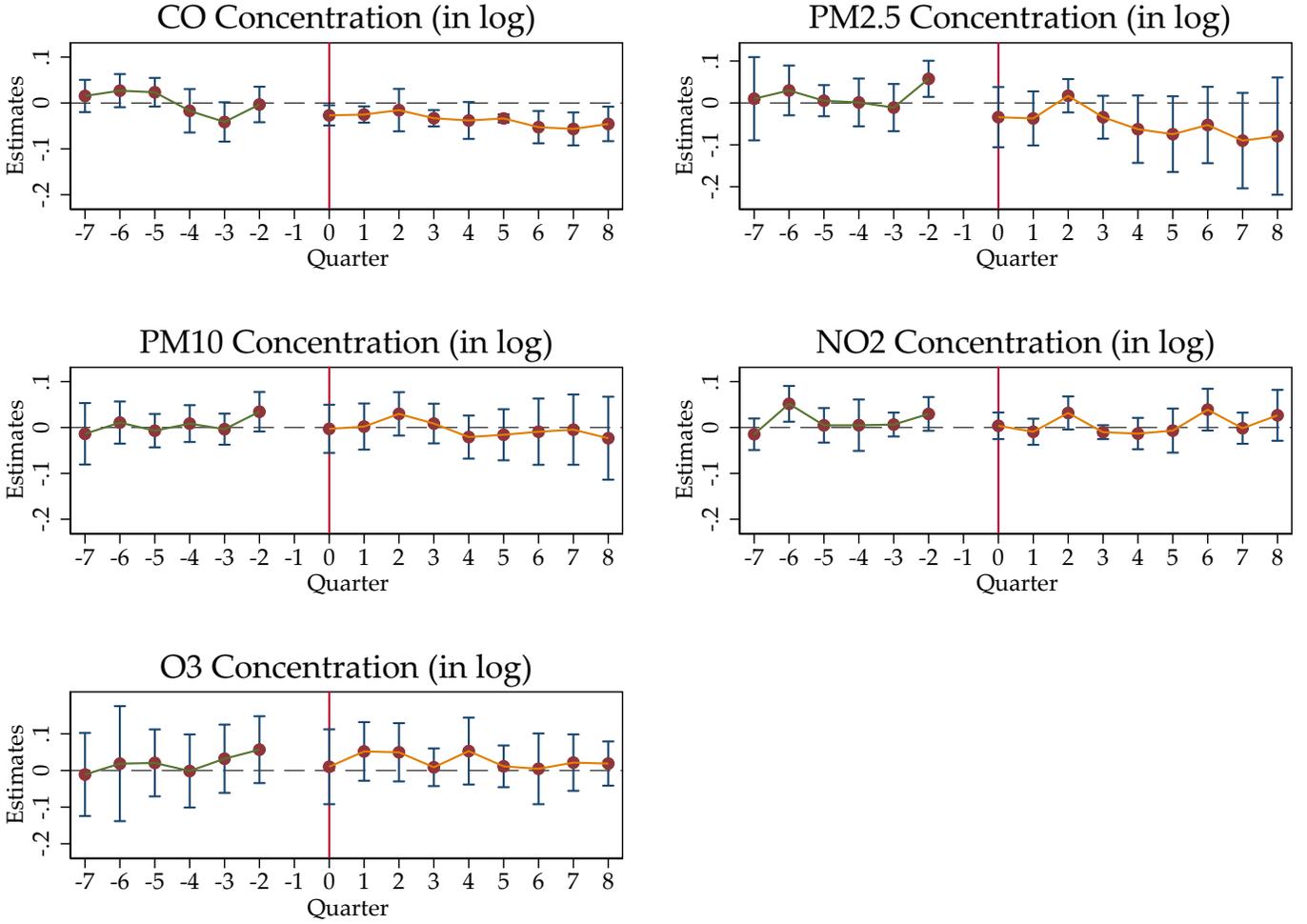


Figure A5: Dynamics of HSR Opening Effects for Air Pollutants along the CSY Highway

Note: We further extend the DID model using an event-study type analysis to compare deweathered air pollution concentrations along the CSY highway two years before and after the HSR opening in the following equation.

$$Air_Pollutant_{c,s,t} = \gamma_t \sum_{t=-7}^8 \times OpeningHSR_t[t \neq -1] \times AQR_{c,s} + \lambda_t + \lambda_c + \epsilon_{c,s,t}.$$

$Air_Pollutant_{c,s,t}$ indicates the natural logarithm of deweathered air pollution concentration reported at AQR grid cell c of road segment s in quarter t . γ_t is a set of year-by-quarter dummies that captures the dynamic effects HSR opening along the CSY highway of our interest, and that, pollution readings one quarter before Chengdu-Chongqing HSR opening was set as the reference category. Hence, $\gamma_t \in \{-7, \dots, -2, 0, 1, 2, \dots, 8\}$. These integers in the set represent 7, ..., 2 quarters prior to the HSR opening, and 1, 2, ..., 8 quarters post the opening, respectively. 0 is the quarter of the policy treatment, which is Q4 2015. λ_t is a set of interactive year-by-quarter fixed effects and λ_c is the AQR grid cell fixed effects. The estimated coefficients of $\gamma_{-7}, \dots, \gamma_{-2}$ before the treatment date present the evidence of the existence of common trends between treated and control observations. The coefficients of $\gamma_1, \dots, \gamma_8$ allow us to access the air pollution levels in the short and long run relative to the reference AQR reading (one quarter before the HSR opening). We found that the insignificant estimates prior to treatment quarter serve as evidence that the control road segments provide a plausible credible counterfactual in the DID design.

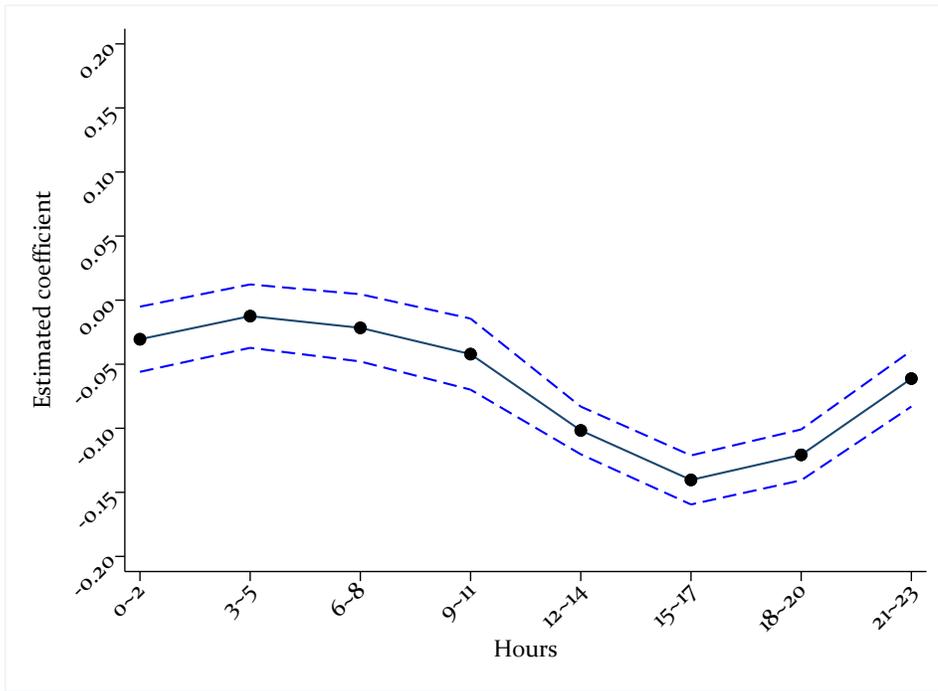


Figure A6: Pollution Concentration for the Cheng-Sui-Yu Highway by Time of the Day (using Stacked Pollutants)

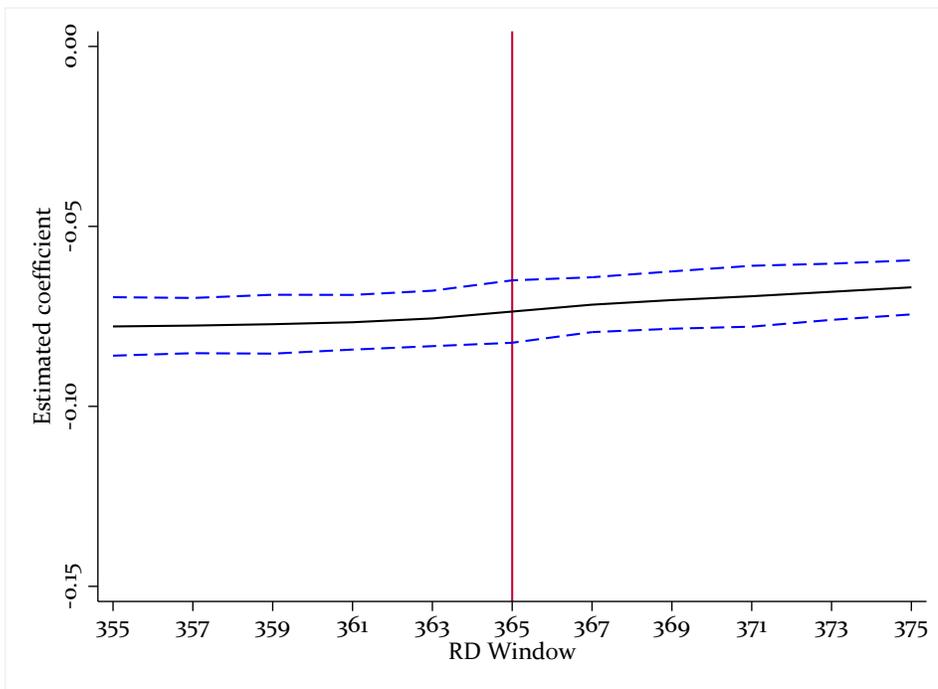


Figure A7: Sensitivity (95% confidence interval) to using Different Bandwidths for the Cheng-Sui-Yu Highway (using Stacked Pollutants)

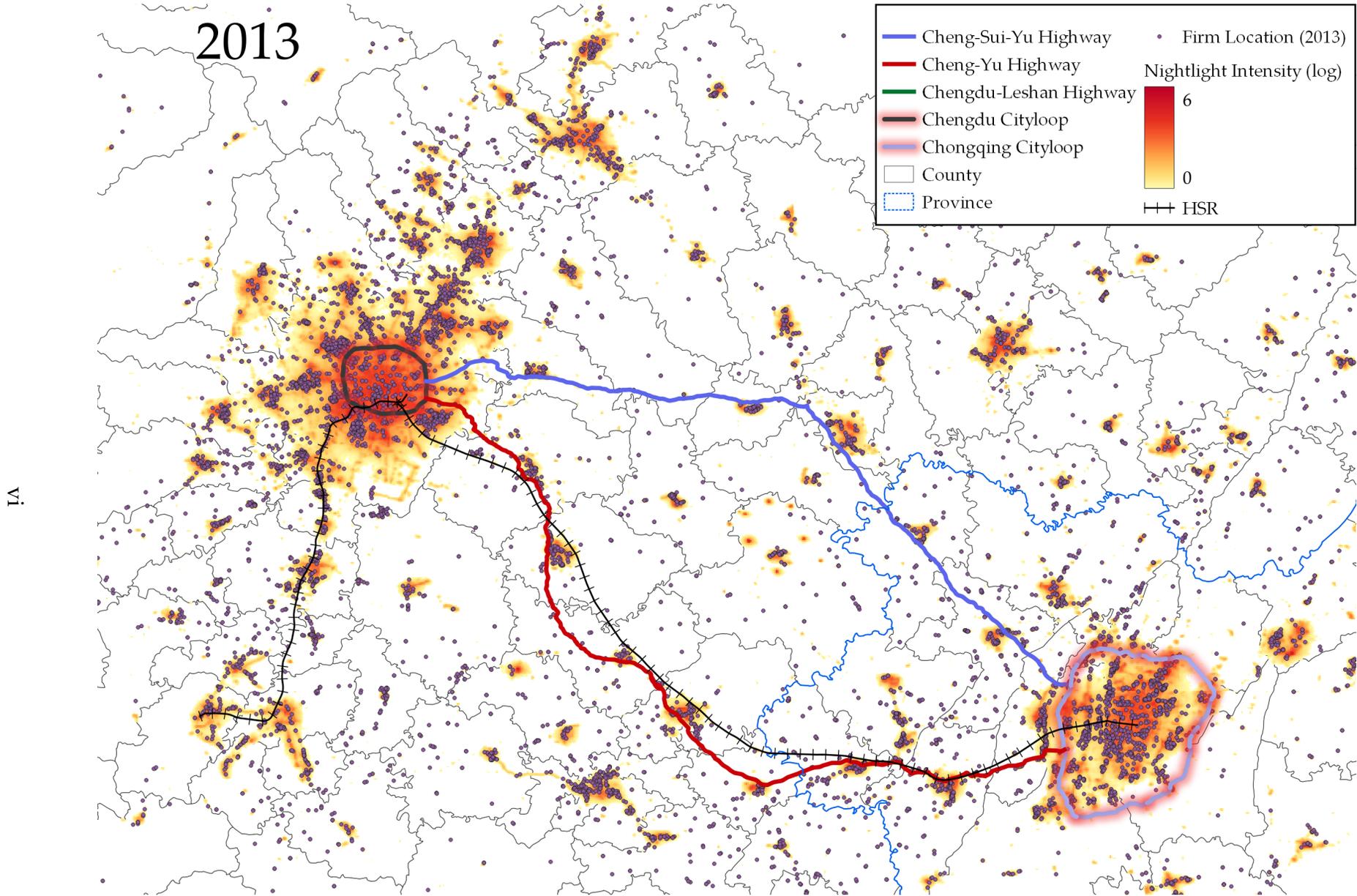


Figure A8: VIIRS Nightlight Intensity and Firm Location in 2013

Note: The stacked firm locations are from 2013 Annual Survey of Industrial Manufacturing Firms. Their exact coordinates are scraped using Baidu Map API. Firm clusters depict a high correlation with nightlight intensity over space.

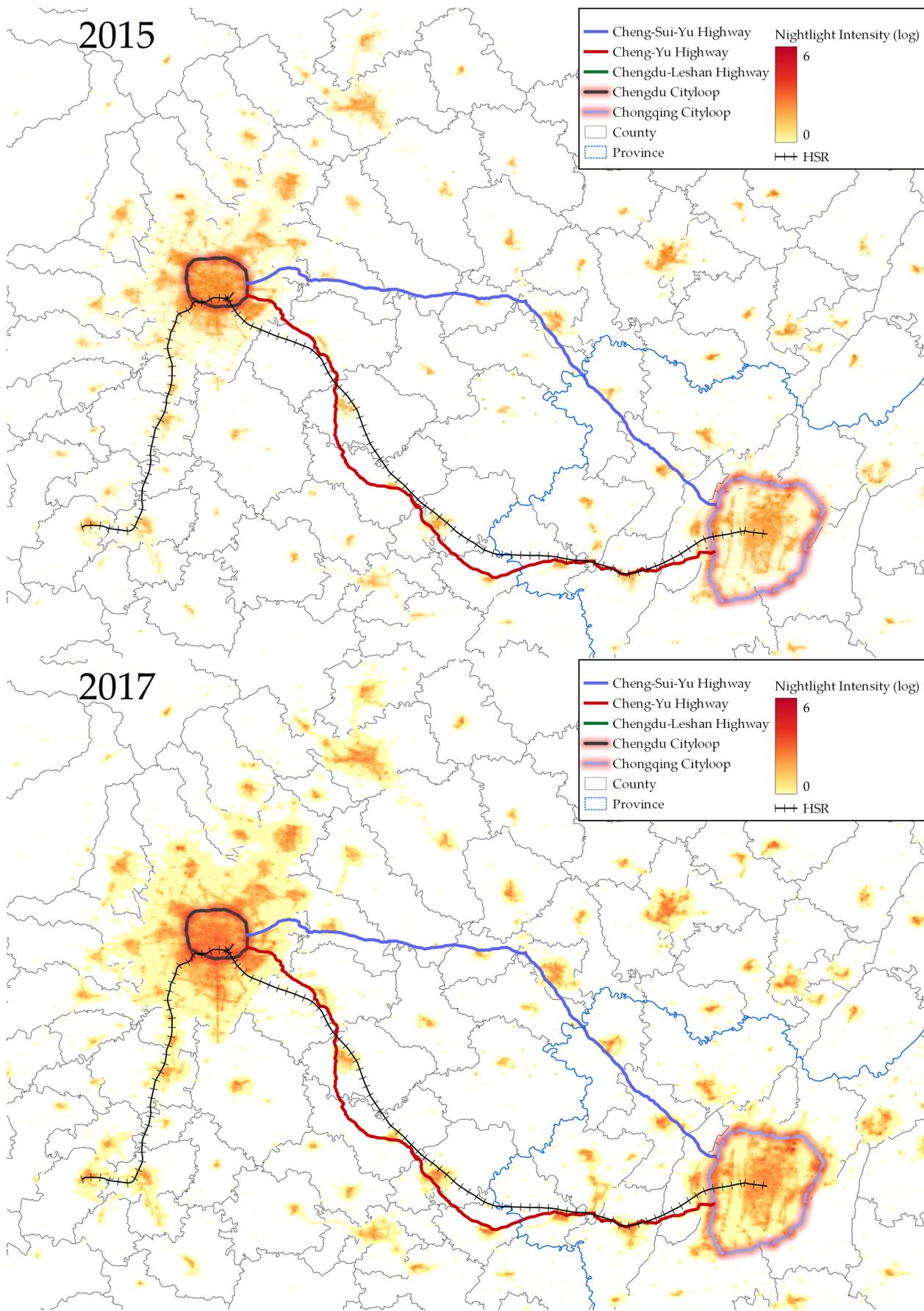


Figure A9: VIIRS Nightlight Intensity over the Period 2015 to 2017

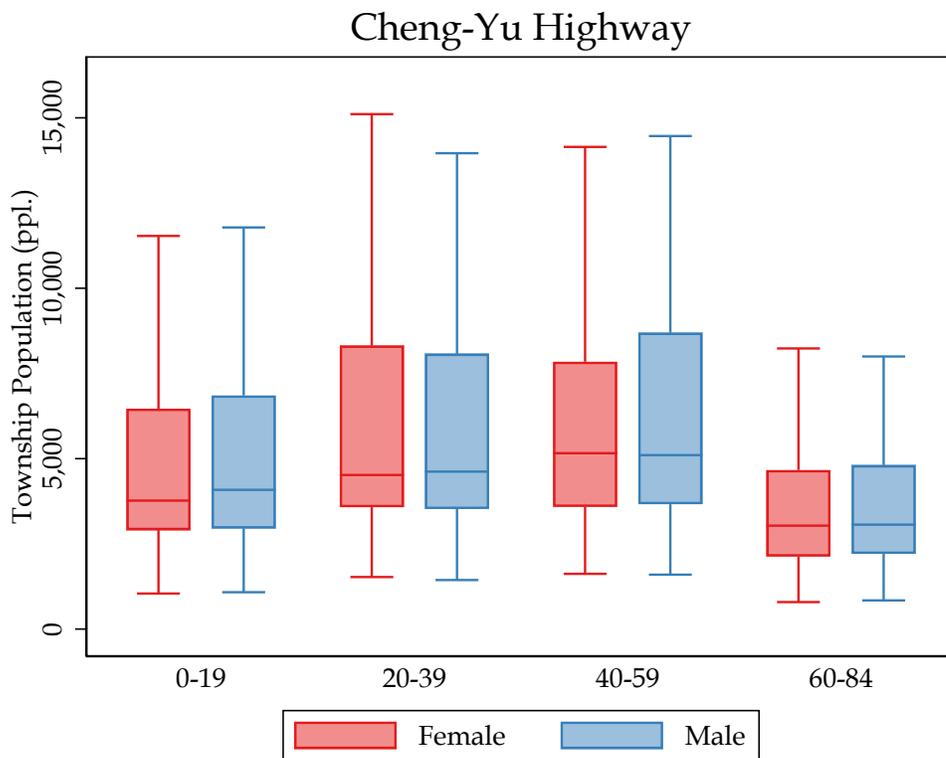
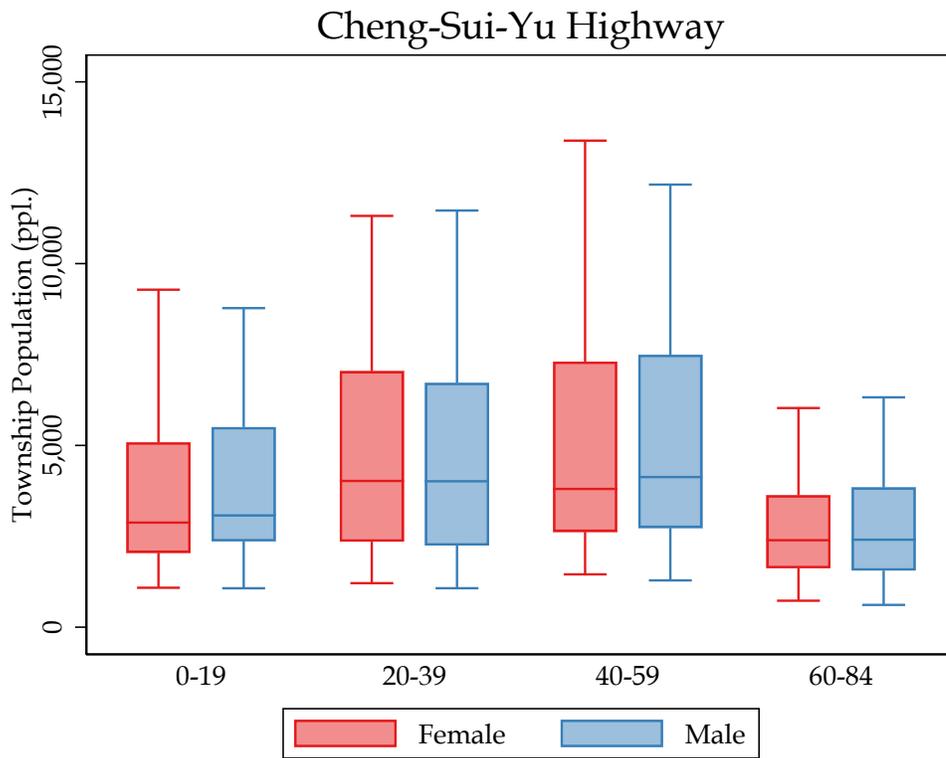


Figure A10: Township Population along the Affected Highways

Table A1: Standard RDiT for Pollution Concentration (by Type of Pollutant) for the Cheng-Sui-Yu Highway

| VARIABLES | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|------------------------------|-----------|-----------|-------------------|-----------------|----------------|------------------|-----------|-----------|-------------------|-----------------|----------------|------------------|
| | Stacked | CO | PM _{2.5} | NO ₂ | O ₃ | PM ₁₀ | Stacked | CO | PM _{2.5} | NO ₂ | O ₃ | PM ₁₀ |
| OpeningHSR | 0.001 | 0.016 | 0.007 | -0.009 | 0.014 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.009* | 0.002 |
| time | (0.002) | (0.011) | (0.007) | (0.008) | (0.011) | (0.005) | (0.001) | (0.001) | (0.002) | (0.002) | (0.005) | (0.002) |
| time ² | -0.000*** | -0.000*** | -0.000* | 0.000 | -0.000*** | -0.000* | -0.000*** | -0.000*** | -0.000 | -0.000** | -0.000*** | -0.000*** |
| time ³ | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| OpeningHSR×time | -0.000*** | -0.000*** | -0.000 | -0.000 | -0.000*** | -0.000 | -0.000*** | -0.000*** | -0.000 | -0.000** | -0.000*** | -0.000** |
| OpeningHSR×time ² | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| OpeningHSR×time ³ | 0.000*** | 0.000** | 0.000** | 0.000 | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000** | 0.000** | 0.000*** | 0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Observations | 175,320 | 35,064 | 35,064 | 35,064 | 35,064 | 35,064 | 175,320 | 35,064 | 35,064 | 35,064 | 35,064 | 35,064 |
| R-squared | 0.982 | 0.916 | 0.971 | 0.960 | 0.972 | 0.979 | 0.993 | 0.994 | 0.996 | 0.978 | 0.993 | 0.995 |
| Data | Raw | Raw | Raw | Raw | Raw | Raw | Deweather | Deweather | Deweather | Deweather | Deweather | Deweather |
| Window (days) | 730 | 730 | 730 | 730 | 730 | 730 | 730 | 730 | 730 | 730 | 730 | 730 |

Note: All AQR grid cell observations are collapsed by highway line. All regressions include pollutant (for stacked data only), hour-by-week FEs and day of Julian year FEs. Time indexes for a third-order polynomial time trends over a four-year window (2014 Jan - 2017 Dec). We also control for lagged outcome variable, and quartics of thermal inversion, precipitation, wind speed, wind direction, relative humidity, air pressure and ground temperature and their one-hour lagged values. Standard errors in parentheses are clustered in five-week intervals. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Augmented RDiT by Type of Pollutants for the Affected Highways (Excluding Observations within Chongqing)

| VARIABLES | Log Residual (Hourly Pollutant Concentration) | | | | | | | | | |
|--------------------------------------|---|--------------------------|------------------------|-----------------------|-------------------------|----------------------|--------------------------|------------------------|-----------------------|--------------------------|
| | (1) CO | (2) PM _{2.5} | (3) NO ₂ | (4) O ₃ | (5) PM ₁₀ | (6) CO | (7) PM _{2.5} | (8) NO ₂ | (9) O ₃ | (10) PM ₁₀ |
| <i>Panel A: Cheng-Sui-Yu Highway</i> | | | | | | | | | | |
| OpeningHSR | -0.127*** (0.006) | -0.330*** (0.012) | -0.123*** (0.011) | -0.018** (0.008) | -0.277*** (0.010) | -0.065*** (0.002) | -0.131*** (0.004) | -0.034*** (0.005) | 0.008*** (0.002) | -0.073*** (0.003) |
| Day | 0.000*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| OpeningHSR × Day | -0.000*** (0.000) | 0.000 (0.000) | -0.000*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.001*** (0.000) | -0.000 (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| <i>Panel B: Cheng-Yu Highway</i> | | | | | | | | | | |
| OpeningHSR | -0.123*** (0.006) | -0.230*** (0.011) | -0.083*** (0.010) | 0.020** (0.009) | -0.200*** (0.010) | -0.069*** (0.002) | -0.064*** (0.004) | 0.002 (0.005) | 0.029*** (0.002) | -0.055*** (0.004) |
| Day | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | -0.000*** (0.000) | 0.000 (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| OpeningHSR × Day | -0.000*** (0.000) | 0.001*** (0.000) | -0.000 (0.000) | 0.001*** (0.000) | 0.001*** (0.000) | 0.000*** (0.000) | 0.001*** (0.000) | 0.000* (0.000) | 0.000*** (0.000) | 0.001*** (0.000) |
| Observations | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 | 17,544 |
| Lag Outcome Var. | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Data | Raw | Raw | Raw | Raw | Raw | Deweather | Deweather | Deweather | Deweather | Deweather |
| Window (days) | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 | 365 |

Note: Regression estimates are obtained from the second stage of the augmented RDiT approach. *OpeningHSR* is the variable of interest. All AQR grid cell observations are collapsed by highway line. Residuals in the second stage are obtained by regressing air pollutants on hour FEs and day of year FEs over a four-year window (2014 Jan - 2017 Dec) in the first stage. In the second stage, we estimate a local linear regression by regressing residuals on lagged air pollutants, the HSR opening date and time trend as a running variable on a narrower two-year window (2015 Jan - 2016 Dec). Regressions using raw data include weather controls in the first stage. 'Raw' indicates non machine-learning processed air pollutant data; 'Deweather' indicates machine-learning processed data. S.E. in parentheses is bootstrapped 500 times. *** p<0.01, ** p<0.05, * p<0.1.