创业与管理学院

School of Entrepreneurship and Management

SHANGHAITECH SEM WORKING PAPER SERIES

No. 2021-002

Natural Gas Transmission Project and Air Pollution in China

Kaiyin Hu

School of Economics, University of New South Wales

Wenhao Wu

School of Entrepreneurship and Management, ShanghaiTech University

April 2, 2021

https://ssrn.com/abstract=3818854

School of Entrepreneurship and Management

ShanghaiTech University

http://sem.shanghaitech.edu.cn

Natural Gas Transmission Project and Air Pollution in China

Kaiyin Hu*[1] and Wenhao Wu†[2]

[1]School of Economics, University of New South Wales

[2]School of Entrepreneurship and Management, ShanghaiTech University

April 6, 2021

Abstract

In this paper, we examine the effect of China's Natural Gas Transmission Project (NGTP) on air pollution. The NGTP is the pipeline network that connects to almost all provinces in China, providing clean energy for industry and households. We use satellite data to measure the country's air quality from 1999 to 2015 at the city level and manually collected information about when each city connected to a natural gas pipeline. We exploit variation in connection time, and use the predicted connection time of pipelines in each city as the instrument to address potential endogeneity concerns. We find that connecting to natural gas pipeline significantly reduces a city's air pollution by 8%.

JEL classification— O1, Q4, Q5

Keywords— Natural Gas Transmission Project, Air pollution, China

1 Introduction

Air pollution is a worldwide environmental problem that causes serious damage to human health. Polluted air usually contains particles and gas that can cause short-term health problems such as respiratory

*kaiyin.hu@unsw.edu.au

†wuwh2@shanghaitech.edu.cn

infections as well as long-term diseases such as lung cancer and heart disease [Pope Iii et al., 2002, Kampa and Castanas, 2008, Neidell, 2004]. However, reducing air pollution and its effects is challenging for two reasons: (1) it is hard to identify its origins,[1] and (2) the pollution itself can be invisible (e.g. carbon dioxide) and long-lasting.[2] Therefore an effective way to control air pollution is to replace "dirty" fuels with green energy in order to provide the necessary energy for economic activities but with less pollution.

China has taken this approach to address its worsening air pollution.[3] Until recently, coal generated over 60% of the country's energy (Figure 1), making China the largest consumer of coal in the world. Since 2011, it has consumed almost 50% of the world's coal (Figure 2). However, there is ample evidence that coal-fueled energy causes serious environmental problems in China [Chen et al., 2012, Kan et al., 2012, Zhang and Smith, 2007]. For example, Almond et al. [2009] show that the coal-fueled winter heating program (which was only implemented in North China) significantly increased the annual average concentrations of total suspended particulates in the winter, compared to South China.

Around 2000, China started to build a national natural gas network to substitute this clean energy for coal. The network is called as natural gas transmission project (NGTP). Since then, total natural gas consumption has increased eightfold; 102 of the country's 293 cities were connected to the national network as of 2015. China has therefore successfully expanded its natural gas facilities, but has this effort actually reduced the country's air pollution? And if so, by how much?

In this paper, we examine NGTP and its impact on the country's air quality from 1999 to 2015. This time range covered the West-to-East Natural Gas Transmission Project Phase I in 2002 to the West-to-East Natural Gas Transmission Project Phase III in 2012. Most main intra-provincial pipelines are constructed during the time periods. Since data on city-level natural gas consumption is not publicly available, we use the date when each city connected to the new gas pipeline to estimate the change in air quality generated by the switch to natural gas. We manually collected information on each city's connection dates from a variety of online sources including local news, media articles, firm memos, and government reports. We merged this data set with satellite data from NASA, which is used to measure air quality. Using the constructed panel of 276 cities from 1999 to 2015, we estimate the causal effect of access to natural gas on local pollution.

There could be endogeneity concerns associated with the connection date because it does not simply indicate when the nearest pipeline passed by the city. To fully connect to a pipeline, it must be extended

[1] It can be generated from smoking, vehicle exhaust, factory emissions, forest fires, etc. [Qin and Oduyemi, 2003] mentioned it can be difficult to pinpoint the precise source of emissions, for instance whether it is natural or anthropogenic.

[2] Carbon dioxide is an example of a type of invisible air pollution [Sorey, 2000]. There are many studies on the impact of long-term exposure to air pollution [Miller et al., 2007, Sastry, 2002].

[3] China's 13th Five-Year Plan (2016–2020) commits to improve air quality and control emissions to enable social development.

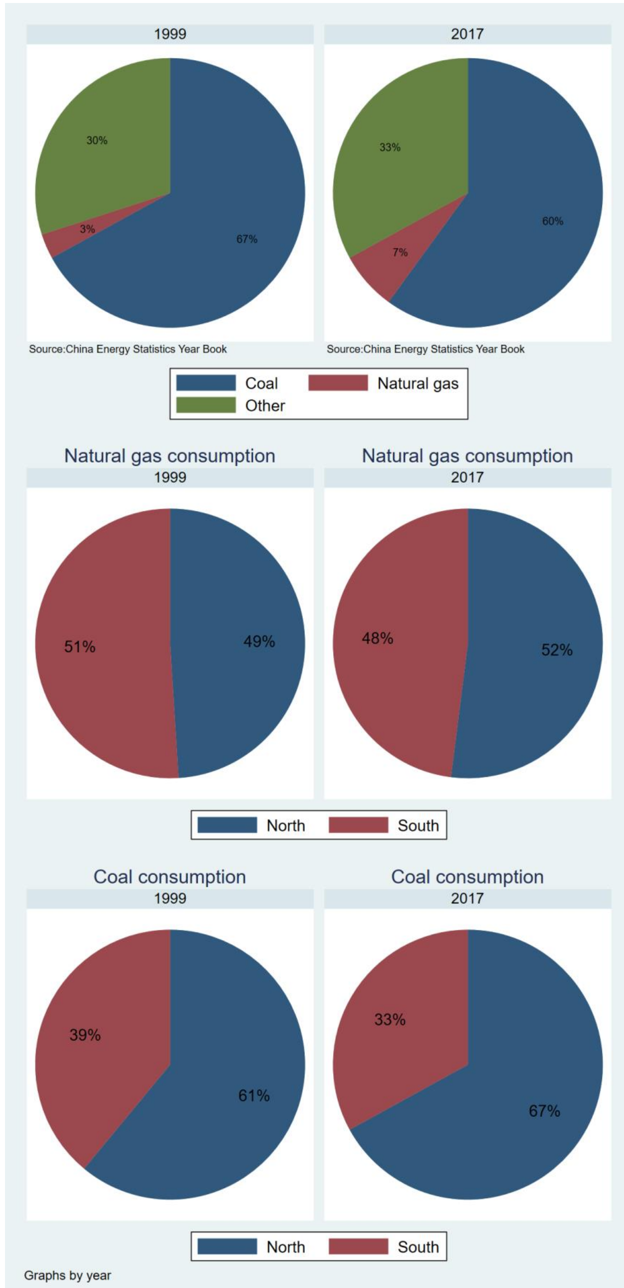

Figure 1: China's Energy Consumption

The first row of Fig. 1 compares China's total coal and natural gas consumption in 1999 vs. 2017. The last two rows illustrate regional energy consumption in 1999 and 2017.
Source: China Energy Statistics Year Book

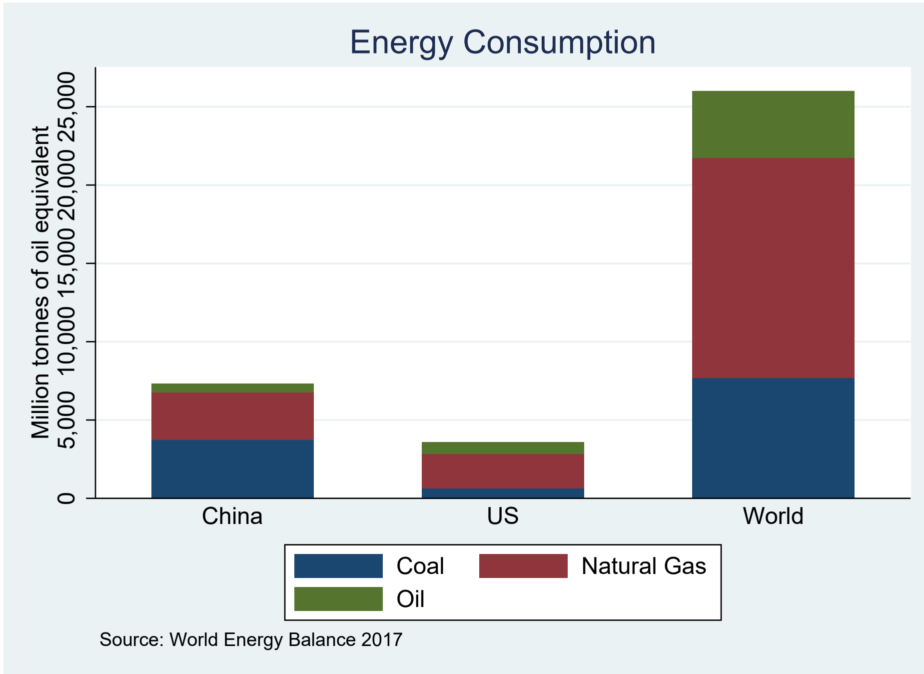

Figure 2: Global Energy Consumption

Source: World Energy Balance 2017

to reach the city and local recipient facilities must be constructed. For example, cities may need to build new pipelines for residential and industrial use, or to upgrade their existing coal–gas pipelines to accommodate natural gas. However, there is large variation in progress in the construction of local recipient facilities. This disparity might signal a local government's attitude toward environmental policy, because local officials ultimately decide when and how quickly to build the facilities. Moreover, cities with more severe air pollution might connect more quickly and/or invest more resources to facilitate this connection. We use the *predicted* pipeline connection time, assuming that pipeline construction proceeds at a constant rate. This variable should be positively correlated with the actual connection time, yet independent of local government actions. Our instrumental variables (IV) regressions therefore provide unbiased estimates of the effect of connecting to natural gas pipelines on local air quality.

We first estimate the impact of connecting to the pipeline on a city's air pollution. According to our two-stage least-squares (2SLS) model, the pipeline connection reduces air pollution by an average of only 8.5%.

We next examine regional and seasonal effects. We find that even though connecting to the pipeline has no substantial differential effect on air pollution in the North vs. South, the growth in air pollution dropped throughout the country after natural gas became available. The seasonal analysis revealed one important difference: in spring, autumn, and winter, connecting to a natural gas pipelines reduces the pollution level and pollution growth rate; yet in the summer, it raises the former and lowers the latter.

We use two instruments to confirm the robustness of our results: (1) the predicted connection time derived from the piecewise line curve between the first and last cities to be connected and (2) the interaction between the predicted connection time and least-cost path. The magnitude of the coefficients is similar using both IVs, which shows that the estimation is robust. We also find that the magnitude of the effect of natural gas increases for cities with shorter least-cost paths to the pipelines. The use of natural gas reduces the growth rate of air pollution as well.

A branch of literature examines the importance of environmental policies in alleviating environmental problems. For example, Nordhaus [2018] points out that if major climate change policies are not implemented, climate change will accelerate over the next century. Alix-Garcia et al. [2015] study a Mexican program in which the federal government pays landowners to protect local forests, and find that it reduces the expected loss of land cover.

Xu and Klaiber [2019] assess how China's natural gas usage in firms affects the country's air pollution. They find that the increase in natural gas intensity and decrease in coal intensity result in a reduction in air pollution. Gao and Zheng [2020] examine nationwide sulfur dioxide (SO2) emissions data from more than 800,000 firm-level observations and find that introducing a natural gas pipeline reduced SO2 discharge intensity in connected cities by 22.8% from 1998 to 2010. They also estimate the heterogeneous effects and find that the reduction effect of natural gas on air pollution increases over time.

Our paper differs from previous research in two respects. First, we have a more complete dataset for estimation. Xu and Klaiber [2019] focus on the effect of one pipeline constructed from 2000 to 2008 and use industry emissions to measure air pollution. Our dataset covers six pipelines and 17 years (1999–2015) and uses satellite data to measure air pollution. Second, we employ a different strategy for our empirical analysis. Xu and Klaiber [2019] and Gao and Zheng [2020] use a difference-in-differences approach to estimate the effect of the natural gas transmission project. Their papers mainly focus on the impact of such policies on either city-level firm emissions or coal intensity. We instead study the policy's impact on air pollution levels in 276 cities. We use an IV to address the potential endogeneity concerns associated with the timing of the pipeline connections. Since pipeline construction is jointly undertaken by national, provincial and local governments, the connection time varies by city. The construction process is complicated, so we find that IV is a better methodology to evaluate the impact of facilitating access to natural gas on levels of air pollution. Our paper is the first to use IV in timing when studying China's dynamic natural gas pipeline policy. We find significant differences between results using ordinary least squares (OLS) and 2SLS.

The remainder of the paper is organized as follows. In Section 2, we introduce the background of

China's natural gas transmission project and the government's decision to construct a pipeline. In Sections 3 and 4, we describe the data set and the empirical strategy. Section 5 reports the results, and Section 6 concludes.

2 Literature Review

Air pollution is a by-product of economic activities that generate negative externalities. It harms people who live near polluted sites, causing diseases such as cough, asthma, lung cancer, etc. This paper examines how stringent environmental regulations can alleviate air pollution.

China's nationwide infrastructure is built to boost regional economic growth and reduce regional inequality. This paper focuses on its nationwide natural gas policy, which is related to the literature on China's infrastructure development and regional inequality. For example, Baum-Snow et al. [2017] use the distance to the transportation network as an IV and find that transportation infrastructure shifts economic activities and populations away from city centers. He et al. [2020b] estimate the effect of China's expressway system on economic growth and air pollution from 2000 to 2012. They use county-level data and discover that counties' connection of local roads with a national expressway have heterogeneous impacts on GDP *per capita*. Banerjee et al. [2020] estimate the long-term effects of a region's transportation network on its economic growth in China. Using the distance to historical transportation links that connect historical Chinese cities to Treaty Ports as the IV, they find close a relationship between cities with a better transportation infrastructure (railroads, coastline, etc.) and positive economic outcomes: counties near the transportation networks have a better transportation infrastructure, leading to faster economic development and greater income inequality. Fan and He [2020] examine the effect of accessing piped drinking water on infant mortality in China. They calculate the least-cost distance between water sources and local infant mortality surveillance areas as an instrument to capture the exogenous variations in the change in the household piped water usage. The increase in the household piped water usage reduces local infant mortality. Particularly in rural areas, households benefit from the piped clean water. Although Chinese government usually sets uniform standards for all places, there has been heterogeneous effects of policies across regions.

This article also contributes to the literature on environmental policies in China. Tanaka [2015] studies the impact of the "Two Control Zones" policies on infant mortality. In 1998, China began shutting down small, inefficient power plants and urging large power plants to reduce their emissions. In the cities in the two control zones, the infant mortality rate fell by 20%; the largest reduction occurred among mothers with a low education and babies in their first month of life. Li et al. [2020] investigate the impact of higher fuel standards on air quality in China. They examine hourly pollution data from 1,492 monitoring stations from 2013 to 2015 and conclude that the higher standards improved air quality. He et al. [2020a] detect an effect of agricultural straw burning pollution on people's health in rural China. They use variations in the number of straw fires near county centers to capture the effect of air pollution on mortality. The main result is that a rise in PM2.5 [4] increases the number of deaths caused by cardiorespiratory diseases.

The effectiveness of environmental policies varies by country. Blundell et al. [2020] study the benefits and costs of the US Environmental Protection Agency's enforcement of the Clean Air Act. Amendments to the act reduced the costs inflicted by air pollution by $35.3 trillion from 1970 to 1990, yet monitoring the regulation cost the government and plants an estimated $831 billion. Moretti and Neidell [2011] estimate the health costs of exposure to ozone in Los Angeles. They use daily boat traffic at the city's port as an instrument to identify the cost of respiratory-related hospitalizations and the effect of avoidance behaviour. Deschenes et al. [2020] analyze the effect of the increase in air pollution on body weight in China. They instrument air pollution growth with thermal inversions and find significant increases in adults' body mass index, overweight and obesity rates when PM2.5 rises within 12 months. Sheldon and Sankaran [2017] examine seasonal haze episodes in Singapore caused by forest burning in Indonesia. The deterioration in air quality during the seasonal haze is associated with increased illnesses such as acute upper respiratory tract infections, asthma, and pneumonia. Greenstone and Hanna [2014] explore the relationship between infant mortality and two air pollution policies in India. They use panel data from 140 cities from 1986 to 2007 and find that air pollution policies are effective at reducing harmful pollution, including NO2, PM10 and SO2. However, they find only a modest correlation between the decline in the infant mortality and air quality policy adoption at the city level.

Few studies have explored the environmental implications of natural gas pipelines in China. Xu and Klaiber [2019] and Gao and Zheng [2020] examine the effect of natural gas usage in each city and find that the usage can reduce firms' coal intensity and SO2 emissions, and increase the intensity of natural gas use. We use a different measure of air pollution levels – satellite-based aerosol optical depth (AOD). Previous studies have not focused on an important feature of Chinese natural gas policy: pipeline construction and connection are jointly decided by central, provincial and local governments, which gives rise to a potential endogeneity problem in the timing of pipeline connections. We use an IV approach for the estimation to address this concern.

[4]PM2.5 is the aerosol particle with aerodynamic diameter less than 2.5 μm involving vehicle emissions, coal burning, etc.

3 Background

3.1 Natural Gas Transmission Project

The NGTP was launched in an effort to balance the uneven geographic distribution of natural gas storage and the disproportionate energy demand in different regions. China's natural gas resources are concentrated in three inland basins in relatively underdeveloped areas: Tarim, Ordos, and Sichuan. Local demand does not support the large-scale exploitation of their gas resources. The project was developed to transport gas to the wealthy eastern coastal provinces, where are far away from the natural gas fields.

The NGTP contains six main pipelines that have nationwide coverage. The best-known pipelines, which run west to east, are known as Phases I, II, and III (abbreviated as WEP-I, WEP-II, and WEP-III). WEP-I was constructed between 2002 and 2004 to transport natural gas from the westernmost Xinjiang Province to Shanghai, the country's largest city. It goes across 10 provinces and municipalities and provides natural gas to 46 cities. WEP-II was built during 2008-2011; construction of WEP-III commenced in 2012 and continues today. These phases also started in Xinjiang Province and ended at the Pearl River Delta (Guangdong Province) and other southeastern provinces. In addition to WEPs, there are another three main pipelines. The Myanmar–China Pipeline (2010–2013) transports natural gas from Myanmar to southwest China, the Sichuan-to-East Pipeline (2007–2010) transports natural gas from the Sichuan Basin to east China, and the Zhong'wu Pipeline (2003–2004) covers cities located in south-central China. Subsidiary pipelines connect the six main pipelines to a broader network (Fig. 3).

3.2 Structure of Government

Inter-province pipelines are built by two state-owned corporations, China National Petroleum and the China Petroleum & Chemical Corporation. The corporations designed the network mainly based on geographic and engineering considerations. After receiving overall approval from the central government, the corporations negotiated with provincial and local governments to allow the pipelines to pass through their territories. The corporations then agreed to supply a certain amount of natural gas to each province.

National energy policy is implemented in a top-down manner that requires coordination between governments at three levels. The central government is committed to building inter-province pipelines that transport natural gas from starting cities to terminal cities. Provincial governments are in charge of building an intra-province network that distributes natural gas from the main route built by the central government. Usually the provincial plan is inclusive and connects all major cities in the province. Local governments must upgrade their energy use facilities to make them natural gas friendly. These upgrades include inner-city pipeline construction and rebuilding residential or industrial energy systems.

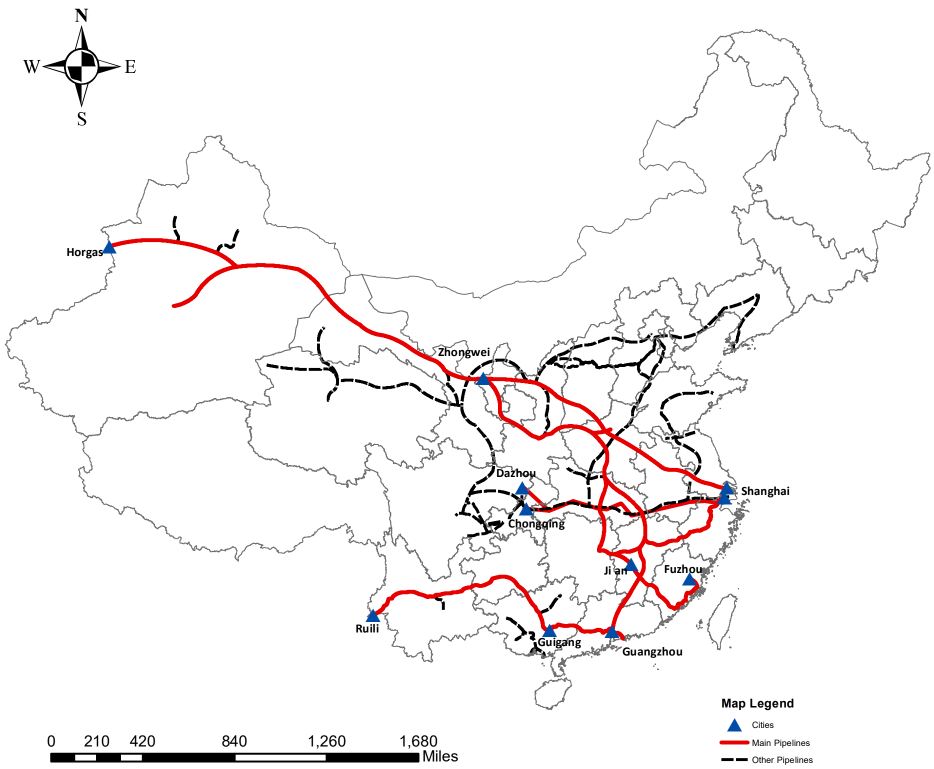

Figure 3: China's Natural Gas Pipeline Network

Blue triangles denote important cities along the six main pipeline routes. WEP-I, WEP-II, WEP-III and the Myanmar–China Pipeline are drawn by Lex Berman from CHGIS, Harvard Dataverse. The Sichuan-to-East and Zhong'wu pipelines are drawn by the authors based on a scanned pipelines network map. The black dashed lines represent other pipelines in China, based on the map drawn by Lex Berman.

Construction at each level is an indispensable link in a city's access to natural gas.

4 Data

We collected data on 276 cities from 1999 to 2015; 37% of these cities received natural gas from main pipelines during this period. The dataset, which is not a balanced panel,[5] consists of four parts:

1. *Satellite-based air pollution data from NASA.* We obtained satellite-based AOD[6] as the measure of PM2.5 from NASA's Modern-Era Retrospective analysis for Research and Applications version 2. The raster data is reported at around 50km*60km latitude by longitude for each month since 1980. High levels of AOD can cause respiratory disease, lung cancer or heart disease. To compute our main dependent variable, city-level air pollution, we use zonal statistics in ArcMap by combining satellite data with China's city-level administrative division map.

2. *Natural gas pipeline network map.* This map is from the ARA Research and Publication Company. It includes the name, geographic location, terminal city, and completion time of each natural gas pipeline in China (completed or in progress). We focus on six pipelines: WEP-I, -II, -III, Myanmar–China, Sichuan-to-East and Zhong'wu. We use geographic distance and data from the Asia South Equidistant Conic project to calculate the centroid of cities. We connect the centroids of cities along the pipelines based on the pipeline construction map to draw a simplified map of the pipeline network using the Arcmap technique (See Fig. 7). The pipelines' locations are approximated by connecting the centroids of the cities. Two types of distance are calculated: the distance from cities to the start point of each pipeline and the distance of cities to each pipeline (from cities to the cities' nearest point on each pipeline).

3. *Cities' connection time.* We manually collected cities' connection times from authorized media and double checked these dates with information from the ARA research company. We searched news[7] with key phrases such as "first utilization time by citizens," "gas supplied by natural gas company," and "natural gas ignition ceremony at terminals." If one city is connected to multiple pipelines, we used its connection to the first pipeline. The blue dots indicate the cities that are connected to the six main pipelines (See Fig. 7).

4. *City statistical yearbooks.* This data set is a panel that covers 290 of China's main 344 cities across 31 provinces and municipalities; it includes cities' demographic and economic characteristics like

[5] We did not drop observations with missing city characteristics data from the sample.
[6] AOD is Aerosol Optical Depth.
[7] All the website links of the news stories were recorded and the screenshots of the web pages were taken.

population, nominal GDP and household energy consumption. We use samples from 1999 to 2015. Some city data are missing before 2003.

We first combine air pollution data set with connection time of the 6 main pipelines. Pipeline indicator represents the pipeline which the city firstly connected to among the six main pipelines. For cities that did not connect to any of the six main pipelines before 2015, we choose the city's nearest pipeline.[8] We add city's demographic and economic characteristics from City Statistical Yearbook to the baseline sample. We exclude the start and terminal cities of each pipeline. They are Bayinguoyu, Horgos, Zhongwei, Dehongdaijingpo, Guigang, Dazhou, Chongqing, Wuhan, Shanghai, Anji, Guangzhou and Fuzhou. The start and terminal cities might cause endogeneity problem if they are in the baseline sample. The pipeline route is designed by the policy makers and constructed by state-owned companies. The start cities own abundant natural resources and the terminal cities are fast growing cities. Cities' GDP, population and other unobservable characteristics, such as the local governments' attitude towards environmental policy and sustainable development, might affect air pollution. Thus we drop those cities from our sample.

We first combine the air pollution data set with the connection times of the six main pipelines. Pipeline indicator represents the pipeline that the city first connected to among the six main pipelines. For cities that did not connect to any of the six main pipelines before 2015, we choose the city's nearest pipeline.[9] We add each city's demographic and economic characteristics from the City Statistical Yearbook to the baseline sample. We exclude the start and terminal cities of each pipeline (Bayinguoyu, Horgos, Zhongwei, Dehongdaijingpo, Guigang, Dazhou, Chongqing, Wuhan, Shanghai, Anji, Guangzhou and Fuzhou) since including them in the baseline sample could cause endogeneity problems. Since cities' GDP, population and other unobservable characteristics – such as local government attitudes towards environmental policy and sustainable development– might affect air pollution, we drop those cities from our sample.

Of the cities in the baseline sample, 102 were connected to the six main pipelines before 2015. The remaining 174 cities are not connected to the main pipelines. [10] We also report our results on selected city samples within different distances with respect to least-cost paths from the main pipelines.

[8]In the baseline sample, some cities did not connect to any main pipeline. Thus in the empirical analysis, we use the nearest pipeline to those cities as the instrument. If the city wants to connect a pipeline and introduce natural gas to the city, the nearest pipeline might be the best choice with least cost.

[9]In the baseline sample, some cities did not connect to any main pipeline. Thus in the empirical analysis, we use the nearest pipeline to those cities as the instrument. If the city decides to connect to a pipeline and introduce natural gas to the city, the nearest pipeline might be the best choice with the least cost.

[10]Natural gas is supplied from resources other than main pipelines in some unconnected cities. For example, coastal cities in Fujian province use imported liquefied natural gas (LNG), and the natural gas used in Sichuan province is mainly supplied by its intra-provincial gas pipelines from local gas fields.

4.1 Summary Statistics

Table 1 displays the summary statistics of the baseline sample. The average air pollution level (AOD) is 0.484[11] and the average air pollution annual growth rate is 1.8% = (0.009 ÷ 0.484). The average distance along the pipelines from the construction start city to the destination city is 3,233 km. The average distance from the centroid city to the pipelines is 400 km.

Table 1: Descriptive Statistics

sample	(1) Mean	(2) Std. Dev.	(3) Min	(4) Max	(5) Obs
Panel A. pipeline connection					
Real time	0.166	0.373	0.000	1.000	4,566
Panel B. environmental characteristics					
Air pollution (AOD)	0.484	0.170	0.052	0.943	4,566
Air pollution growth (ΔAOD)	0.009	0.048	-0.188	0.188	4,289
Distance along the pipeline (km)	3,253.535	1,558.030	0.000	5,113.877	4,566
Distance to the pipeline (km)	510.017	511.131	0.207	2,257.409	4,566
Panel C.demographic characteristics					
City population (10,000 persons)	419.829	318.558	14.550	3,941.000	4,559
City population density (persons/sq.km)	415.444	351.172	4.700	11,564.000	4,558
Panel D.economic characteristics					
GDP (10,000,000,000 RMB)	10.797	16.783	0.169	256.691	4,492
Primary	1.079	1.004	0.004	15.545	4,423
Secondary	5.390	7.813	0.034	94.953	4,424
Tertiary	4.452	9.306	0.036	205.943	4,421
Per capita GDP (RMB)	28,069.320	27,616.773	99.000	467,749.000	4,050
Industry firm (number)	945.046	1,297.676	19	12,491	4,548
Industrial SO2 emissions (ton)	57,787.321	62,621.707	2.000	831,372.000	3,521
Industrial dust emissions (ton)	31,792.023	122,438.182	34.000	5,168,812.000	3,518

Notes: Variables are observed at the city and year levels. *Real time* is a dummy variable that indicates whether a city connected to the main pipeline in a given year. The baseline sample uses data from the China City Statistical Yearbooks (1999–2015). Distance was computed by the authors using Arcmap.

Table 2 describes the differences between cities that are connected to the main pipelines and those that are not. While the average air pollution *level* is higher in the former, the air pollution *growth* is not significantly different. Connected cities are located closer to construction cities and the pipelines, and have significantly larger populations and GDP devoted to both the agricultural and industrial sectors. Average GDP per capita is not significantly different. The average number of industrial firms, average SO2 emissions in industry and average dust emissions are higher in cities that are connected to the pipelines. As the characteristics of connected and disconnected cities are different, it is difficult to find a control group for connected cities in the sample.

[11] Kong et al. [2016] study the correlation between AOD and PM2.5 in Beijing. Using the linear regression mentioned in their paper, the average AOD is in the range of light pollution. According to China's Environmental Protection Administration, the 24-hour average PM2.5 in light pollution standard is 75-115 $\mu g\ m^3$. Figure 6 depicts the average AOD from 1999 to 2015 in mainland China.

Table 2: Descriptive Statistics of Groups

Sample	Connected Cities			Disconnected Cities			P Value
	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs	
Panel A. pipeline connection							
Real time	0.434	0.496	1,752	0.000	0.000	2,814	0.000
Panel B. environmental characteristics							
Air pollution (AOD)	0.540	0.169	1,752	0.450	0.162	2,814	0.000
Air pollution growth (AOD)	0.010	0.048	1,647	0.008	0.048	2,642	0.183
Real distance along the pipeline (km)	3,044.386	1,576.421	1,752	3,383.753	1,532.394	2,814	0.000
Real distance to the pipeline (km)	307.389	292.683	1,752	636.174	573.721	2,814	0.000
Panel C.demographic characteristics							
City population (10,000 persons)	443.294	258.791	1,749	405.225	349.903	2,810	0.000
City population density (persons/sq.km)	464.225	280.981	1,751	385.014	385.522	2,807	0.000
Panel D.economic characteristics							
GDP (10,000,000,000 RMB)	11.271	15.747	1,726	10.501	17.394	2,766	0.135
Primary	1.148	0.961	1,703	1.036	1.027	2,720	0.000
Secondary	5.662	7.547	1,703	5.219	7.972	2,721	0.067
Tertiary	4.594	8.097	1,701	4.362	9.988	2,720	0.421
Per capita GDP (RMB)	28,167.950	28,076.859	1,563	28,007.334	27,329.160	2,487	0.857
Number of industrial firms	1,071.689	1,316.284	1,744	866.277	1,279.895	2,804	0.000
Industrial SO2 emissions (ton)	53,286.722	43,155.119	1,358	60,612.940	72,076.376	2,163	0.001
Industrial dust emissions (ton)	23,337.992	19,624.932	1,357	37,100.733	155,223.246	2,161	0.001

Notes: Variables are observed at the city and year levels. The baseline sample uses data from the China City Statistical Yearbooks (1999–2015). P-values of t-statistic on the difference in means between cities that are connected to the main pipelines and those that are not are reported. Distance was computed by the authors using Arcmap.

5 Empirical Strategy

To capture the causal effect of pipeline connection on air pollution, we need to address the endogeneity problem in the OLS estimation. Unobserved confounding factors could be driving the bias in the simple correlation between pipeline connection and city-level pollution. First, air pollution could be highly correlated with economic activities. Fast-growing cities are more likely to connect to the main natural gas pipelines at an earlier stage and to also experience high levels of air pollution. This could upwardly bias the coefficient between pipeline connection and air pollution. Second, if environmental policies are implemented more strictly in big cities, the coefficient between pipeline connection time and air pollution could be downward biased. To overcome both challenges and identify the causal effect, we need a random source of variations for air pollution.

Previous studies have treated natural gas usage or pipeline connection time as exogenous to the individual or firm, perhaps because individuals and firms cannot decide whether to use natural gas or not. If a natural gas pipeline is built in the region, the coal gas pipeline will eventually stop being used to supply coal gas to households. Nor can individuals or firms quickly respond to short-term changes in air pollution. However, the pipeline connection time could be correlated with air pollution. Furthermore, the pipeline connection time or natural gas usage time might be correlated with unobserved confounding factors such as economic activities, environmental policies, etc. Another reason that we do not consider pipeline connection time to be exogeneous to air pollution is that connected cities have higher primary and secondary GDP (GDP from the agricultural or industrial sectors) and more firms than unconnected cities (see Table 2). In this paper, we use the predicted connection time as the source of random variations in long-term air pollution, as explained in the next section.

5.1 Predicted Connection Time

The variations in cities' connection times are based on provincial and local governments' decisions: more polluted cities are more concerned about their environment and more likely to connect to the pipelines at an earlier stage. From this process we can see that there could be large variations in local connection dates despite there being common dates when the national pipeline becomes available in an area. The actual dates are affected by the building of national pipelines, but are also determined by policies implemented by provincial and local governments. This may raise an endogeneity problem in the connection time, leading to an upward bias in the measurement of the effect of pipeline connection on air pollution. To address this problem, we use the predicted connection time as an instrument. The predicted connection time is the *hypothetical* year in which a main pipeline is available to a city, after fixing the starting and ending years of the entire pipelines and assuming that the construction continues at a constant rate. The predicted connection time is the earliest time a city could connect to the pipeline, regardless of whether it does so. The predicted connection time is only correlated with the progression of the construction of national pipelines; it is independent of variations in provincial and local governments' policy making. The gap between predicted connection time and real connection time reflects whether the local government is eager to introduce natural gas to the city.

In this section, we introduce how we calculate the predicted connection time based on the simplified real pipelines we draw from scratch with ArcGIS (see Fig 7). Initially, we split each real pipeline into bins of equal length according to how many years it takes to finish the construction. Predicted year is tagged to the bins whose construction is supposed to be completed, provided that the construction proceeds at a constant rate. Then, we use the Arcmap technique to draw a line from each city perpendicular to a real pipeline, intersecting at one point. The year associated with the bin where the point falls is the predicted connection time.

The construction of WEP-I and WEP-II started at multiple cities at the same time but were completed in different years, as shown in Table 3. We partition WEP-II into two segments: Horgos to Zhong'wei (west segment), and Zhong'wei to Guang'zhou (east segment). Similarly, WEP-III is partitioned into three segments: Horgos to Zhong'wei (west segment), Zhong'wei to An'ji (middle segment) and An'ji to Fu'zhou (east segment). Since the middle segment is in progress before 2015, we do not consider this segment in the pipeline.

Figure 4 presents the nearest points to each sample city and how to define the predicted connection

Table 3: Construction Periods of Pipelines

Pipeline Name	Commence Time	Completion Time
West-to-East Natural Gas Transmission Project Phase I	Jul-2002	Oct-2004
WEP Phase II Horgos to Zhong'wei (west segment)	Feb-2008	Dec-2009
WEP Phase II Zhong'wei to Guang'zhou (east segment)	Feb-2008	Jun-2011
WEP Phase III Horgos to Zhong'wei (west segment)	Oct-2012	Aug-2014
WEP Phase III Zhong'wei to An'ji (middle segment)	Oct-2012	In progress
WEP Phase III An'ji to Fu'zhou (east segment)	Oct-2012	Dec-2016
Myanmar-China Natural Gas Pipeline	Sep-2010	Aug-2013
Sichuan-to-East China Pipeline	Aug-2007	Mar-2010
Zhong'wu Pipeline	Aug-2003	Nov-2004

Notes: All the information is collected from authorized media and double checked with the partial information available from the ARA research company.

time. We draw the real pipeline that connects Horgos, Zhongwei, and Guangzhou. As the real construction periods are from 2008 to 2009 and from 2008 to 2011, we split the two segments into two even bins and four even bins, respectively. We label the bins of first segment from 2008 to 2009 and those from the second segment from 2008 to 2011. We draw a line starting from the centroid of one city and extend it perpendicular to the nearest pipeline. The city's predicted connection is the year associated with the bin that corresponds to the intersection point. In the graph, the predicted connection time for Lanzhou is 2009 and 2010 for Wuhan. If a city cannot be projected onto the pipeline, the predicted connection time for the city is either the first or last year of construction, depending on whether it is near the starting point or ending point of the pipeline.

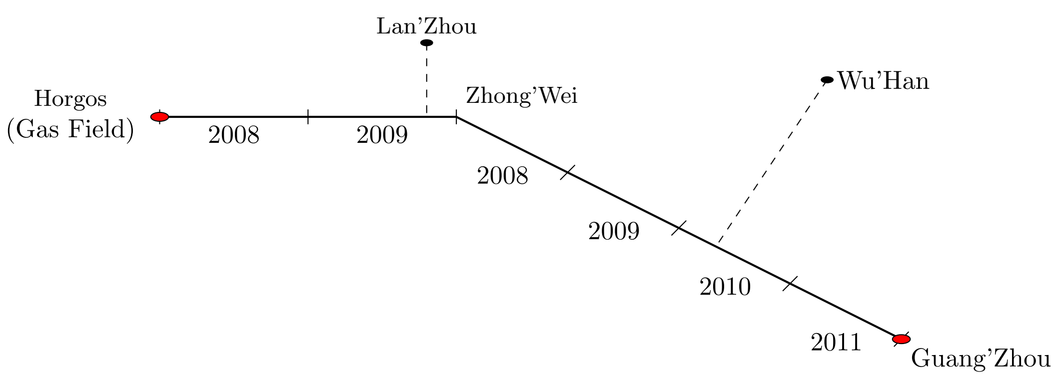

Figure 4: Predicted Connection Time

Fig. 4 is a real pipeline that connects Horgos, Zhongwei, and Guangzhou. We split the pipeline segments into bins based on the construction period. We draw a perpendicular line from the centroid of the city to the nearest pipeline. The city's predicted connection time is the year associated with the bin where the perpendicular line passes through.

For cities that do not connect to any of the six main pipelines, we use the predicted connection time of the nearest pipelines. Thus our main source of plausibly exogenous variations in real connection time stems from the predicted connection time.

5.2 Main Estimation

To estimate the average effect of a natural gas pipeline connection on air pollution, we use predicted connection time as our IV in the 2SLS model. Eq. (1) describes the first stage, which examines the correlation between the predicted connection time and the real connection time. The second stage is specified by Eq. (2) to estimate the effect of real connection time on air pollution.

$$\mathbf{1}\{real\ time\}_{it} = \delta \mathbf{1}\{predicted\ time\}_{it} + x_{it}\Gamma + \rho_i + \rho_t + \epsilon_{1it} \tag{1}$$

$$Pollution_{it} = \phi \mathbf{1}\{\widehat{real\ time}\}_{it} + x_{it}\Pi + \theta_i + \theta_t + \epsilon_{2it} \tag{2}$$

In the model, $\mathbf{1}\{real\ time\}_{it}$ is a dummy variable that indicates whether city i began to use natural gas supplied by the main pipelines in year t. For each city, $\mathbf{1}\{real\ time\}_{it}$ equals 1 if a city is connected to the pipelines in year t, and 0 otherwise. In other words, the time indicator is an approximation of whether citizens in city i are using natural gas provided by the main pipelines in year t. The real connection time indicator of disconnected cities equals 0 for all periods. $\mathbf{1}\{predicted\ time\}_{it}$ is a dummy variable that indicates whether the city is predicted to be connected in period t. If the city's predicted connection time is period t, it is more likely to connect to the real pipeline after time t. $Pollution_{it}$ is the annual average AOD, which measures the level of air pollution in city i in year t. The models include both city fixed effects and year fixed effects to control for invariant city characteristics and year shocks. Standard errors are clustered at the city level.

We use x_{it} as a vector of specifications. The baseline specifications are log population and log real GDP, which might be correlated with the change in air pollution if population and GDP are changed by the natural gas transmission project. For example, people are more likely to migrate to big cities, where the air pollution level increases sharply with the explosion of the population and economic growth. However, the effect of pipeline connection on cities might be small, as natural gas consumption is low in China: (1) coal generated 60% of the country's total energy needs in 2015 and (2) natural gas generated only 3% of the country's total electricity in 2017. We do not use geographic distance from cities to the real pipelines as a baseline specification in the regression because the effect of geographic distance is absorbed by the city fixed effects. We also drop start and terminal cities in the sample as they are usually big cities and the location might be endogenous to the pipeline routes.

We expect the coefficient of interest δ in the first stage to be positive as a city is more likely to have a real connection to a pipeline if it is predicted to have a connection to the pipeline at that time. The OLS model captures the average effect of pipeline connection on air pollution. After natural gas is introduced to the cities through pipelines, the number of firms and households that use natural gas increases slowly. Therefore China's natural gas consumption is increasing steadily. Thus, we first estimate the average impact of connecting to the pipeline on air pollution in the OLS model. Then, the predicted connection time is used as the IV in the second stage. In the 2SLS regression, ϕ is the effect of the real pipeline connection on the city's air pollution. The natural gas pipeline network provides cities with clean energy to substitute for coal, thus we expect the sign of ϕ to be negative. Note that we restrict our attention to the effect of the initial connection to the main pipelines on air pollution.[12]

5.3 IV Validity

This section tests the validity of the instrument. We use a zero-stage equation that is similar to the zero-stage model used in Sequeira et al. [2019]. We examine the effects of one-year-lagged characteristics of connected cities on the real pipeline connection year. We add the interaction between pipeline fixed effects and province fixed effects to control for variations between provinces and pipelines. We expect the effect to be significant. If so, the year of the pipeline connection is associated with a city's pre-connection characteristics. Then we evaluate the effects of one-year-lagged characteristics of connected cities on the pipeline's predicted connection year, which is calculated using the pipeline's construction period and the ArcGIS method as described above. We expect the effect to be insignificant, which would provide supportive evidence that the IV is exogeneous to city characteristics before the pipeline connection. Finally, we exploit the effects of one-year-lagged city characteristics on the gap between real connection year and predicted connection year. We expect the effect to be significant, which implies that a city's waiting time for a pipeline connection is correlated with its pre-connection characteristics.

The zero-stage equation for cities that have connected to a pipeline is as follows:

$$\{year\ of\ connection\}_i = \eta x_{year-1,i} + \alpha_{pipeline} + \alpha_{pipeline} \times \beta_{province} + \epsilon_i \tag{3}$$

$\{year\ of\ connection\}_i$ is the year of real connection time; $\alpha_{pipeline}$ is the province fixed effects and $\alpha_{pipeline} \times \beta_{province}$ is the interaction between the pipeline fixed effects and province fixed effects. $x_{year-1,i}$ is the one-year-lagged cross-sectional city characteristic before the real connection time. η measures the correlation between cities' characteristics in the pre-connection period and real pipeline connection

[12] In the sample, some cities have multiple connection times because they are connected to both WEP-I and WEP-II or to both the Sichuan-to-East and Zhong'wu pipelines. Cities also used natural gas from other sources prior to the main pipelines in two situations: (1) they were supplied natural gas by pipelines and gas fields located in the same province or (2) the natural gas was supplied by other sources, such as LNG imported from overseas. We do not include natural gas supplied by intra-provincial pipelines or other sources in our estimation because the effects are likely to be relatively small and hard to detect.

time. Our model includes air pollution, population density, GDP, number of industrial firms, and gross industrial output as measurements of cities' environmental wellbeing and economic growth. We expect η to be negative because more polluted or faster-growing cities are more likely to connect to the pipeline at an earlier stage.

As the data sample includes cities that connect to the six main pipelines, we use pipeline fixed effects to capture variations in the connection time of each pipeline. We also use the interaction of pipeline fixed effects and provincial fixed effects to capture variations in each pipeline's province connection time. To test the validity of the instrument, we perform the same regression on the year of predicted connection time. We do not expect the predicted connection year to be correlated with $x_{year-1,i}$ because the predicted connection year is exogenous to the lagged cross-sectional city characteristics. To further examine the correlation between the real connection year and city characteristics, we perform the same regression on the duration of the connection. The duration is the gap between the predicted connection year and the real connection year. Since city characteristics may affect both the real connection year and the duration, we expect the sign of η to be the same as the sign of the coefficient of real connection time on city characteristics.

Table 4 reports the results of zero-stage cross-sectional estimations. Our sample only includes 103 connection cities. Both route fixed effects and the interaction between route fixed effects and province fixed effects are included in all models with the full sample. Column 1 is the full sample excluding three cities that do not have air pollution data. Column 2 is the sub-sample of cities connected to WEP-I and only the province fixed effect is included. Column 3 is the sub-sample of cities connected to WEP-I and the Zhong'wu Line (the "sub-pipeline" of WEP-I).

Table 4: Zero-stage Cross-section Estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Real Connection Year						
Lagged Variable(x)	Total	WEP-I	WEP-I & Zhongwu				
AOD mean	-5.242	-31.54**	-20.95*				
(AOD)	(0.244)	(0.045)	(0.071)				
Population Density				-0.00204*			
(person/sq.km)				(0.092)			
City GDP					-4.74e-08*		
(100 million yuan)					(0.066)		
Number of Industrial Firms						-0.000492**	
(unit)						(0.023)	
Gross Industrial Output Value							-2.35e-08
(current price)							(0.102)
Constant	2011.4***	2021.4***	2016.3***	2009.8***	2009.1***	2009.0***	2008.8***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	100	43	56	103	103	100	100
R-squared	0.795	0.179	0.154	0.798	0.796	0.784	0.784
Route fixed effects	YES	NO	NO	YES	YES	YES	YES
Province fixed effects	NO	YES	NO	NO	NO	NO	NO
Route fixed effects * Province fixed effects	YES	NO	YES	YES	YES	YES	YES

Notes: Table 4 shows the zero-stage cross-sectional results using the sub-sample of connected cities. All regressions control for route fixed effects and the interaction between route fixed effects and province fixed effects. In the model, we use robust standard errors. * p<0.10, ** p<0.05, *** p<0.01.

The air pollution level has no significant effect on each city's pipeline connection time in the total sample. Applying the same regression to sub-samples, we find that the effect of the lagged air pollution level among the cities that connected to WEP-I is significant at the 1% level and is larger in magnitude. The difference between the two coefficients might be attributable to two causes. (1) The first main inter-provincial pipeline, WEP-I, is connected to 43% of the cities in the total sample, which can capture clear variations in connection time raised by air pollution in the pre-connection period. (2) The air pollution level is small and the year of connection cannot precisely pin down the change in both timing and air pollution.

The population density, city GDP, number of industrial firms and gross industrial output value are negatively correlated with the year of connection, which implies that cities that have a higher GDP and rely more on industry are more likely to use natural gas at an earlier time. This result can be explained as the difference in cities' attitudes towards environmental protection and ability to upgrade their local natural gas facilities in a short time.

To further demonstrate the validity of the instrument, we perform the same regression on the instrument, i.e., the year of predicted connection. Table 5 demonstrates that air pollution and city characteristics have no significant effect on the predicted connection time. This is evidence that the IV is not correlated with cities' air quality or economic development. As for the regression on the duration [13] between real and predicted connection time in Table 6, we find the results are quite similar to the one for the true connection year, which also supports our hypothesis that cities' connection times vary with economic growth and environmental conditions.

[13]Only for the dependent variable is duration which is calculated as (Real Connection time - Predicted Connection Year+2) because (Real Connection time - Predicted Connection Year) equals -1 for some cities that connected to WEP-I.

Table 5: IV Validity Estimation: Predicted Connection Time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Predicted Connection Year						
Lag Variable(x)	Total	WEP-I	WEP-I & Zhongwu				
AOD mean	0.521	0.212	0.162				
(AOD)	(0.103)	(0.241)	(0.242)				
Population Density				0.0000660			
(person/sq.km)				(0.184)			
City GDP					1.61e-09		
(100 million yuan)					(0.160)		
Number of Industrial Firms						0.00000384	
(unit)						(0.254)	
Gross Industrial Output Value							5.15e-10
(current price)							(0.177)
Constant	2005.9***	2003.8***	2003.8***	2006.2***	2006.1***	2006.2***	2006.2***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	237	133	157	241	237	239	239
R-squared	0.997	0.761	0.895	0.997	0.997	0.998	0.997
Route fixed effects	YES	NO	NO	YES	YES	YES	YES
Province fixed effects	NO	YES	NO	NO	NO	NO	NO
Route fixed effects * Province fixed effects	YES	NO	YES	YES	YES	YES	YES

Notes: Table 5 shows the zero-stage cross-sectional results for the regressions using the sub-sample of connected cities. All regressions control for route fixed effects and the interaction between route fixed effects and province fixed effects. We include robust standard errors. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 6: IV Validity Estimation: Duration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Duration between Real Connection Year and Predicted Connection Year						
Lagged Variable(x)	Total	WEP-I	WEP-I & Zhongwu				
AOD mean	-6.037	-31.54**	-0.282				
(AOD)	(0.198)	(0.045)	(0.935)				
Population Density				-0.00211*			
(person/sq.km)				(0.082)			
City GDP					-5.03e-08*		
(100 million yuan)					(0.065)		
Number of Industrial Firms						-0.000506**	
(unit)						(0.020)	
Gross Industrial Output Value							-2.39e-08
(current price)							(0.102)
Constant	6.975***	19.44**	4.088**	4.934***	4.275***	4.332***	4.156***
	(0.004)	(0.013)	(0.022)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	100	43	57	103	103	100	100
R-squared	0.118	0.179	0.355	0.118	0.110	0.116	0.0998
Route fixed effects	YES	NO	NO	YES	YES	YES	YES
Province fixed effects	NO	YES	NO	NO	NO	NO	NO
Route fixed effects * Province fixed effects	YES	NO	YES	YES	YES	YES	YES

Notes: Table 6 shows the cross-sectional results for the regressions using the sub-sample of connected cities. The dependent variable is duration, which is calculated as (Real Connection time - Predicted Connection Year+2) because (Connection time - Predicted Connection Year) equals -1 for some cities that connected to WEP-I. All regressions control for route fixed effects and the interaction between route fixed effects and province fixed effects. We include robust standard errors. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

6 Results

6.1 First-Stage Results

Table 7 reports the first-stage results. It shows that the predicted connection time is strongly correlated with the real connection time. Standard errors are clustered by 276 cities in the model, which covers over two-thirds of China's cities. The models control for both city and year fixed effects. The predicted connection time has a positive impact on the real connection time. If the indicator of predicted connection time moves from 0 to 1, it increases the average possibility of connecting to the real pipeline by 0.245. The predicted connection time can explain 60% of the variations in the real time.

Table 7: First Stage: The Effect of Predicted Connection Time on Real Time

	(1)	(2)	(3)	(4)
	Real Time			
Predicted Connection Time	0.245***	0.244***	0.247***	0.246***
	(10.407)	(10.360)	(10.473)	(10.467)
Log real GDP		-0.003		-0.040
		(-0.068)		(-0.766)
Log city population			0.132**	0.159**
			(2.266)	(2.558)
Constant	0.033**	0.070	-0.737**	-0.463
	(2.556)	(0.128)	(-2.165)	(-0.835)
Observations	4,566	4,492	4,559	4,490
R-square	0.62	0.61	0.62	0.61
F statistics	101.7	100.7	103.0	102.8
City fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES

Notes: Table 7 shows the first-stage results for the regressions using the full sample. All regressions control for city and year fixed effects. Standard errors are clustered by 276 cities. Predicted connection time and distance were computed by the authors using Arcmap. F-statistics in the first stage are above 100. * p<0.05, ** p<0.01, *** p<0.001.

We use log real GDP and log city population as baseline specifications. Columns 2–4 show that log real GDP does not have a significant effect on the real time, but that log city population does have a positive effect on real time. Including both log real GDP and log population in the regression yields similar coefficients for predicted connection time. This implies that the estimation is robust to using different control variables.

6.2 2SLS Results

To capture the NGTP's effect on air pollution, we use the predicted connection time as the IV to resolve the biased estimation because the real pipeline connection is endogenous to city-level air pollution. Table 8 presents the regression estimates from Eq. (2) with various specifications. Column 1 reports the OLS results using both log real GDP and log population as baseline specifications. Columns 2–5 show the estimations of the 2SLS model, and we use different specifications in the model.

Table 8: 2SLS: The Effect of Real Connection Time on Air Pollution Level

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV
	Air pollution (annual averaged AOD)				
Real time	0.023***	-0.041***	-0.042***	-0.041***	-0.041***
	(6.634)	(-2.628)	(-2.699)	(-2.673)	(-2.674)
Log real GDP	-0.036***		-0.033***		-0.037***
	(-4.405)		(-3.631)		(-3.793)
Log population	0.009			-0.004	0.017
	(0.789)			(-0.394)	(1.575)
Constant	0.815***				
	(8.354)				
Observations	4,490	4,566	4,492	4,559	4,490
R-square	0.96				
Kleibergen-Paap Wald F statistics		101.7	100.7	103.0	102.8
City fixed effects	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES

Notes: Table 8 shows the OLS and second-stage results using the full sample. All models control for city and year fixed effects. Standard errors are clustered by 276 cities. Predicted connection time was computed by the authors using Arcmap. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

The OLS results indicate that, conditional on baseline specifications, cities' connecting to the pipelines leads to a 0.023 increase in air pollution. This result is contrary to our expectations and might be due to the endogeneity problem. For example, cities may be motivated to upgrade their gas facilities at an earlier stage if they experience worse air pollution. We are concerned about endogeneity in the connection time and therefore use the predicted connection time as an instrument to resolve this problem.

Columns 2–5 show the results using IV; both the sign and magnitude of the coefficients of real time change. Column 2 is the model without any specification. The effect on air pollution becomes significantly negative, indicating that the NGTP reduces city-level air pollution. The average air pollution level is 0.484 in Table 1, and the NGTP reduces air pollution by 8.5% on average. After controlling for log real GDP and log population, the coefficients of real time on air pollution do not change much. The consistency of the coefficients implies that the IV is exogenous to log real GDP and log population.

6.3 Alternative Instruments

To further specify the endogeneity in cities' connection and connection time, we use three new IVs to explain the impact of natural gas connection and economic growth on air pollution. The three new IVs are the interaction of real connection time and log city characteristics (log population density, log real GDP, and log total value of industrial firm). Those instruments measure the effect of both pipeline connection and cities' economic characteristics after connection. The effects can be explained as: (1) Cities with a higher potential natural gas demand are more likely to engage in pipeline construction and gas facility upgrading. For example, gas companies prefer cities with large populations and a potential green energy market. (2) Cities with a higher GDP are more capable of building sub-pipelines and upgrading their facilities, which shortens the construction duration. (3) Cities prefer clean energy if they are more concerned about air pollution. Under the central government's restricted pollution regulation, cities are under more pressure to use green energy if they have higher industrial GDP.

Table 9 shows the results after using the interaction as instruments. The first stage results show that all the cities' characteristics interacting with predicted connection time have positive effects on the real connection time. These results imply that cities' population density, real GDP and industrial output value increase the probability of real pipeline connection if the predicted pipeline connection equals 1.

In the 2SLS estimation, we find that the real pipeline connection is linked to a reduction in air pollution if we use real GDP or gross industrial output value as the IV. This result proves that connecting to the pipeline has negative effects on air pollution, which may be because cities with a higher GDP or industrial value are more motivated to connect to the pipeline at an earlier stage, which in turn reduces their air pollution. There is no effect on air pollution if we use the interaction between predicted connection time and population density. Household natural gas consumption is smaller than industrial natural gas consumption, thus we do not observe an effect of population density on air pollution.

7 Heterogeneous Effects

In this section, we test two heterogeneous effects of pipeline connection on air pollution. First, we examine the heterogeneous *regional* effects (results reported in Table 10). To identify these effects, we classify cities into four groups: Municipalities, northern cities, southern cities and other. The municipalities are Beijing and Tianjin, which did not connect to the main inter-provincial pipelines. As these two municipalities connected to other pipelines and did so earlier than 2004, we exclude them from the sample and find that the effect of pipeline connection on air pollution changes little in either sign and magnitude. The other cities are Tibet, Qinghai and Hainan. Tibet and Qinghai have vast landscapes at high altitude,

Table 9: The Effect of Cities Characteristics after Pipeline Connection on Air Pollution

	(1)	(2)	(3)
	Air pollution (annual average AOD)		
	OLS		
Real Time	0.0253***	0.0278***	0.0279***
	(0.004)	(0.004)	(0.004)
	Reduced		
Predicted Connection Time	0.0006	-0.0007*	-0.0004*
× Log City Characteristics	(0.001)	(0.000)	(0.000)
	2SLS		
Real Time	0.0124	-0.0279*	-0.0247*
	(0.012)	(0.014)	(0.014)
	First Stage		
Predicted Connection Time	0.0479***		
× Log Population Density	(0.004)		
Predicted Connection Time		0.0241***	
× Log real GDP		(0.002)	
Predicted Connection Time			0.0174***
× Log Gross Industrial Output Value			(0.002)
Observations	4,641	4,573	4,573
R-squared	0.624	0.616	0.619
Kleibergen-Paap Wald F statistics	104.3	107.9	114.3
City fixed effect	YES	YES	YES
Year fixed effect	YES	YES	YES

Notes: Table 9 shows the effect of city characteristics on air pollution after pipeline connection. All regressions control for city and year fixed effects. Standard errors are clustered by 276 cities. Predicted connection time was computed by the authors using Arcmap. All F-statistics of interaction terms in the first stage are above 70. * p<0.10, ** p<0.05, *** p<0.01.

and Hainan is the largest island in China. We exclude all three cities due to their unique geographic characteristics; Column 2 demonstrates that the effect on air pollution remains negative. The rest of the cities are grouped based on the Qinling-huaihe line. The northern cities are located north of this line and have had free heating in the winter since the 1960s; southern cities are below this line and do not have free heating. The 2SLS estimations show that the effects of pipeline connection on air pollution in North and South China are small and insignificant. One potential explanation might be that the effect is hard to detect in such a small sample.

Table 10: Heterogeneous Effects: By Region

	(1)	(2)	(3)	(4)
	Air Pollution Level (AOD)			
	Exclude municipalities	Exclude other	North	South
	OLS			
Real Time	0.0287***	0.0273***	0.0185**	0.0282***
	(0.004)	(0.004)	(0.009)	(0.004)
	Reduced			
Predicted Connection Time	-0.0101***	-0.0109***	-0.000335	-0.000343
	(0.003)	(0.003)	(0.006)	(0.004)
	2SLS			
Real Time	-0.0407***	-0.0436***	-0.000919	-0.00141
	(0.015)	(0.015)	(0.017)	(0.016)
	First Stage			
Predicted Connection Time	0.247***	0.249***	0.365***	0.244***
	(0.024)	(0.024)	(0.039)	(0.031)
Observations	4,456	4,410	1,548	2,828
R-squared	0.613	0.614	0.634	0.614
Kleibergen-Paap Wald F statistics	102.8	101.2	80.72	57.91
City fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES

Notes: Table 10 shows the regional heterogeneous effects of pipeline connection on air pollution. Log GDP and log population are used as control variables in the model. All regressions control for city and year fixed effects. Standard errors are clustered by cities. Predicted connection time was computed by the authors using Arcmap. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Second, we assess heterogeneous *seasonal* effects. We collect seasonal connection time data using year and month information. As shown in Table 11, having a pipeline connection decreases the air pollution level in winter (Dec-Feb), spring (Mar-May) and autumn (Sep-Nov), but increases it in summer (Jun-Aug). This could be because industrial and household energy usage peaks in the summer, leading to an energy shortfall that is filled by using coal or coal gas instead, which sharply increases the pollution from the emissions. In Table 18, being connected to a pipeline decreases the pollution growth in winter, spring and summer.

Table 11: Heterogeneous Effects: By Season

	(1)	(2)	(3)	(4)
	Air Pollution Level (AOD)			
	Winter	Spring	Summer	Autumn
	OLS			
Real Time	0.0536***	0.0163***	0.0198***	0.0277***
	(0.006)	(0.005)	(0.005)	(0.005)
	Reduced			
Predicted Connection Time	-0.0102**	-0.0225***	0.0164***	-0.0248***
	(0.004)	(0.004)	(0.003)	(0.006)
	2SLS			
Real Time	-0.0417**	-0.122***	0.0785***	-0.110***
	(0.019)	(0.027)	(0.017)	(0.030)
	First Stage			
Predicted Connection Time	0.245***	0.240***	0.263***	0.276***
	(0.024)	(0.026)	(0.027)	(0.027)
Observations	4,566	4,566	4,566	4,566
R-squared	0.615	0.394	0.402	0.411
Kleibergen-Paap Wald F statistics	101.7	78.40	88.09	96.96
City fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES

Notes: Table 11 shows the seasonal heterogeneous effects of pipeline connection. We find that the signs of the coefficients change little if we add control variables in the models. All regressions control for city and year fixed effects. Standard errors are clustered by cities. Predicted connection time was computed by the authors using Arcmap. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

8 Robustness Check

8.1 Instrument — Piecewise Linear Curve

As a robustness check, we use a new IV — predicted connection time calculated from piecewise linear curve in the 2SLS model. We draw piecewise linear curves between the start and terminal cities as straight pipeline segments. Those curves are green lines in Fig. 7 between important cities. We split each line by construction periods in Arcmap and calculate each city's predicted connection time using the piecewise linear curve. This changes 26 cities' predicted connection time.

Table 12 shows the 2SLS results using the piecewise linear curve. The real connection time's effect on air pollution is around -0.03, which is similar to the effect estimated as -0.045 in the model using real pipeline as the IV. The results are robust if we add GDP and population as control variables in the regressions.

Table 12: IV: Straight Lines

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV
Real Time	0.0235***	-0.0328**	-0.0344**	-0.0331**	-0.0336**
	(0.004)	(0.015)	(0.015)	(0.015)	(0.015)
Log real GDP	-0.0356***		-0.0335***		-0.0372***
	(0.008)		(0.009)		(0.010)
Log Population	0.00879			-0.00507	0.0165
	(0.011)			(0.011)	(0.011)
Constant	0.815***				
	(0.098)				
Observations	4,490	4,566	4,492	4,559	4,490
R-squared	0.959				
Kleibergen-Paap Wald F statistics		107.5	107.5	107.5	107.5
City fixed effect	YES	YES	YES	YES	YES
Year fixed effect	YES	YES	YES	YES	YES

Notes: Table 12 shows the results using the predicted connection time from piecewise linear curve as the IV. All models control for city and year fixed effects. Standard errors are clustered by 244 cities. Predicted connection time was computed by the authors using Arcmap. * p<0.05, ** p<0.01, *** p<0.001

8.2 Instrument — Least-cost Path

We use least-cost path as a new IV to conduct another robustness check. This measures both the time and money spent on pipeline construction, because it takes more time and money to build a sub-pipeline to connect to the main pipeline if the land is steep or mountainous. Fan and He [2020] study the impact of clean piped water on infant mortality, using the least-cost path from local water sources to infant mortality surveillance sites as an instrument to resolve the endogeneity associated with piped water usage.

In this section, we use Arcgis to draw the least-cost path from city centroids to natural gas pipelines. We first collected landscape data from the SRTM 90m DEM Digital Elevation Database, which is released by NASA and distributed by the US Geological Survey. We calculated and classified the landscape slope produced from DEM data into nine groups. The larger the value of the landscape slope, the higher the path cost. Then we used the cost distance toolbox in Arcmap to draw the least-cost path from city centroids to pipelines. Higher path costs imply more potential time and money spent during the pipeline construction. Figure 5 is the landscape slope map of mainland China; the purple routes represent the least-cost paths from city centroids to pipelines.

Figure 5: Least-Cost Paths from City Centroids to Pipelines

Fig. 5 is a landscape slope map of mainland China. The blue dots represent the city centroids. The purple routes are the least-cost paths from city centroids to the pipelines. The least-cost path was computed by the authors using Arcmap.

Since the least-cost path is collinear with city fixed effects, we use the interaction between predicted connection time and least-cost path as the IV. Both predicted connection time and least-cost path are exogenous to air pollution. A higher least-cost path means more time and money was spent on the sub-pipeline. The correlation between the interaction and real connection time demonstrates the effect of the predicted total construction time on the real connection time. We expect that the predicted connection time has a positive effect (and the least-cost path has a negative effect) on real connection time. Thus, the combined effect of the interaction on air pollution is ambiguous.

Table 13 shows the results. We use log(least-cost path)+10 to make sure the value of the least-cost path is normal distributed and above 0.[14] We use cities with different least-cost paths as sub-samples. In the first stage, the correlation between interaction and real connection time is positive. This implies that the effect of predicted connection time has more power than the effect of least-cost path on the real connection time. The 2SLS results show that the effect of pipeline connection on air pollution is significantly negative and the magnitude is around 0.05, which is similar to the previous results. The reduction effect on air pollution becomes larger if the cities have a lower least-cost path to the main pipelines in the sub-samples .

Table 13: IV: Least-cost Path

	(1)	(2)	(3)
	Air Pollution Level (AOD)		
	Total Sample	GEM Distance 0-10	GEM Distance 0-8
	OLS		
Real Time	0.0301***	0.0304***	0.0293***
	(0.004)	(0.004)	(0.004)
	Reduced		
Predicted Connection Time	-0.00126***	-0.00129***	-0.00142***
$\times$ ln (Least Path Cost)+10	(0.000)	(0.000)	(0.000)
	2SLS		
Real Time	-0.0555***	-0.0577***	-0.0680***
	(0.020)	(0.021)	(0.023)
	First Stage		
Predicted Connection Time	0.0231***	0.0227***	0.0213***
$\times$ ln (Least Path Cost)+10	(0.003)	(0.003)	(0.003)
Observations	4,196	4,094	3,973
R-squared	0.942		
Kleibergen-Paap Wald F statistics	65.50	60.55	52.72
City fixed effects	YES	YES	YES
Year fixed effects	YES	YES	YES

Notes: Table 13 shows the results using the distance IV: the interaction between predicted connection time and least-cost path. The least-cost path is calculated as ln (Least-Cost Path)+10. All the models control for city and year fixed effects. Standard errors are clustered by 244 cities. Predicted connection time and least-cost path were computed by the authors using Arcmap. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

[14] The value of log least-cost path is between -8.8 and 3.01.

8.3 Sample Selection

We use different least-cost distance and different time span as sub-samples to check the robustness of the estimations. Table 14 shows the 2SLS results. It shows that the magnitudes of the effects increase if we narrow the time span. The results imply that real pipeline connection has a larger impact on cities with lower least-cost paths. The results are insignificant if we include only connected cities, which may be because the difference between connected cities and unconnected cities is large and cannot be neglected.

8.4 Fixed Effects

To better capture within-province variations in connecting to pipelines, we add the interaction between route fixed effects and province fixed effects in the model. Table 15 shows that the estimations do not change compared to Table 8 and Table 16. The results imply that city and time fixed effects have already captured within-province variations.

9 Air Pollution Growth

In this section, we use the change in the air pollution level as the dependent variable and the same vector specifications as in Eq. (1). Both city and year fixed effects are included. The 2SLS model is as follows:

$$\Delta Pollution_{it} = Pollution_{it} - Pollution_{i,t-1} = \kappa \mathbf{1}\{real\ time\}_{it} + x_{it}\Lambda + \chi_i + \chi_t + \epsilon_{3it} \tag{4}$$

Air pollution growth is calculated as the air pollution level at period t minus the air pollution level at period $t-1$. We expect that the pipeline connection slows down the air pollution growth as well. Table 16 shows the results of the effects of connecting to the pipeline on air pollution growth. The pipeline connection can reduce air pollution growth by -0.07. The estimated effect is negative, and the magnitude of the OLS estimate is smaller than the 2SLS estimate. Table 17 and 18 demonstrate the regional and seasonal heterogeneous effects of connecting to the pipeline on air pollution growth. We find that the 2SLS estimations change little across different sub-samples: connecting to the pipeline decreases the air pollution growth. Table 19 reports the results of the robustness check using route fixed effects interacted with province fixed effects. The results of air pollution growth are also robust to these new fixed effects.

10 Conclusion

China has been constructing a nationwide network of natural gas pipelines for the past 30 years to transport natural gas from resource-rich cities to fast-growing cities. In this paper, we estimate the effect of the NGTP on air pollution in China. We use exogenous variances in predicted connection time to capture the effect of using natural gas on air pollution. The predicted connection time is determined by the joint decision of central, provincial and local governments. We find that on average, the NGTP has substantially reduced air pollution. Connecting to the pipeline can reduce a city's air pollution by 8%. We find significant seasonal heterogeneous effects and insignificant modest regional effects on air pollution. We also calculate the least-cost path and find that the results are also robust if we use the interaction between least-cost path and predicted connection time as the IV. Using least-cost path as the IV allows us to more precisely predict the real connection time, because higher least-cost paths indicate that more time and money are spent on pipeline construction. In the future, we can address two important questions related to pipeline construction and air pollution. First, we can use information about subsidiary pipelines to develop a broader picture of the effects of switching from coal to natural gas on air pollution. We can estimate how the environment can impact pipeline construction and how local policies affect industrial and residential energy usage. Second, we can explore how households' natural gas usage affects family members' health. As women and children are more likely to be exposed to harmful indoor pollution, such a change in household energy consumption might improve their health.

11 Figures

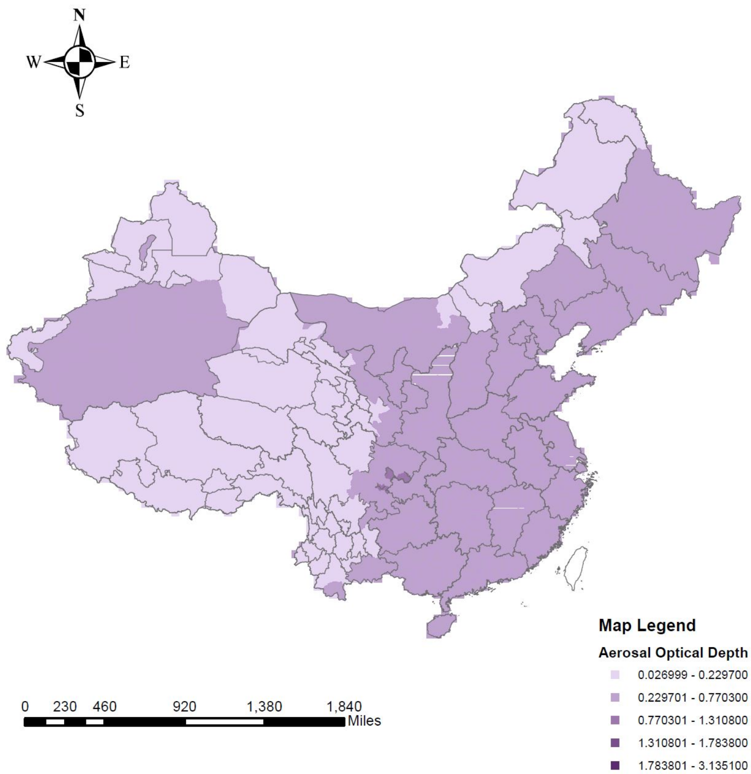

Figure 6: Average City-level AOD, 1999–2015

Fig. 6 depicts average air pollution in mainland China from 1999 to 2015. Kong et al. [2016] studied the correlation between AOD and PM2.5 in urban Beijing. To simplify the calculation of the correlation in different regions of China, I use the correlation in urban Beijing in the paper to roughly describe air pollution, which is PM2.5$= 74 * AOD + 18$. According to the China Environmental Protection Administration, the 24-hour average PM2.5 ranges are: optimal 0 $\mu g/m^3 - 35$ $\mu g/m^3$; good 35 $\mu g/m^3 - 75$ $\mu g/m^3$; light pollution 75 $\mu g/m^3 - 115$ $\mu g/m^3$; moderate pollution 115 $\mu g/m^3 - 150$ $\mu g/m^3$; high pollution 150 $\mu g/m^3 - 250$ $\mu g/m^3$; severe pollution 250 $\mu g/m^3 - 500$ $\mu g/m^3$. The AOD value is classified into five groups with respect to PM2.5 ranges: 0.0269–0.2297, 0.2297–0.7703, 0.7703–1.3108, 1.3108–1.7838, 1.7838–3.1351.

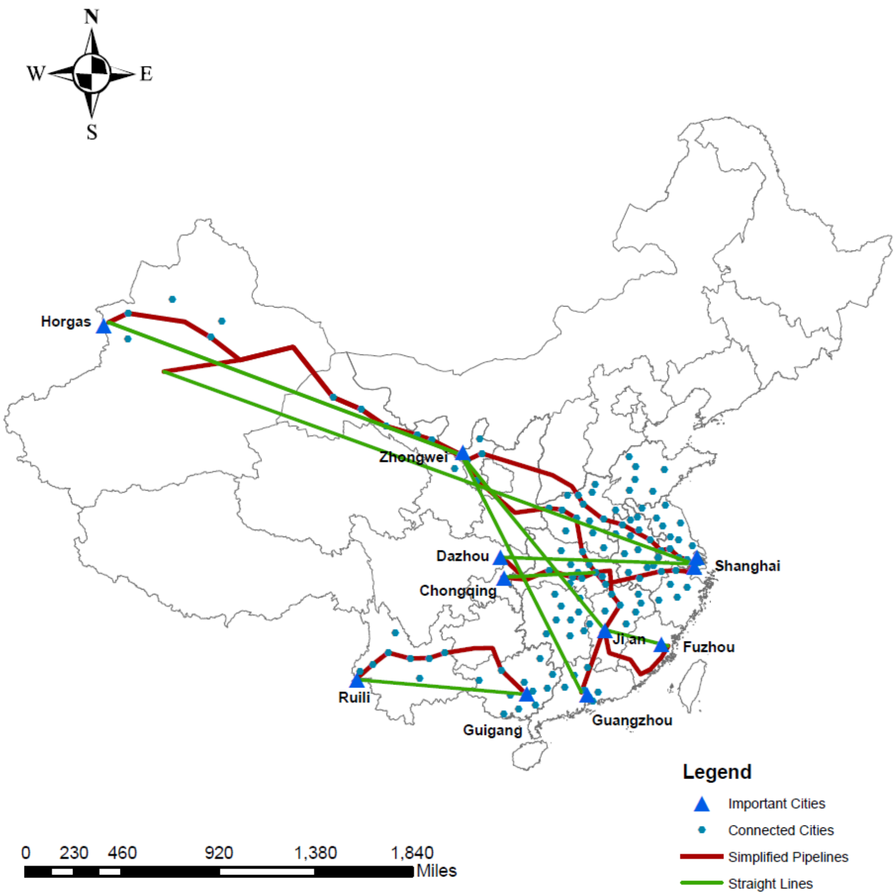

Figure 7: Simplified Pipeline Network

Fig 7 shows the simplified pipeline network drawn by the authors based on the China pipeline network map. The six main pipelines (WEP-I, -II, -III, Myanmar-China, Sichuan-to-East and Zhong'wu) are marked in red. All six main pipelines were constructed between 2000 and 2015. Green lines are piecewise linear curves connecting the important cities. Important cities on the six main pipelines are marked as blue triangles. The blue dots represent cities that are connected to the six main pipelines. Cities from Sichuan province are not included on the map because they all used natural gas before the 1990s.

12 Tables

Table 14: 2SLS: with Sample Selection

	(1)	(2)	(3)
	1999-2015	2001-2015	2003-2015
	Only Connected Cities		
Real Time	0.0121	0.000612	-0.0135
	(0.009)	(0.008)	(0.009)
	Total Sample		
Real Time	-0.0481***	-0.0580***	-0.0755***
	(0.017)	(0.015)	(0.015)
	GEM Distance 0-10		
Real Time	-0.0481***	-0.0580***	-0.0755***
	(0.017)	(0.015)	(0.015)
	GEM Distance 0-8		
Real Time	-0.0504***	-0.0608***	-0.0788***
	(0.018)	(0.016)	(0.017)

Notes: Table 14 shows the second-stage results for the regressions with selected samples using predicted connection time. The sub-samples are only treated cities, least-cost path less than 15, 10 and 8. All regressions control for city and year fixed effects. Standard errors are clustered by cities. Predicted connection time and least cost path were computed by the authors using Arcmap. * p<0.05, ** p<0.01, *** p<0.001.

Table 15: 2SLS: Route Fixed Effects

	(1)	(2)	(3)	(4)
	Air Pollution Level (AOD)			
	Exclude municipality	Exclude other	North	South
	OLS			
Real Time	0.0293***	0.0278***	0.0177**	0.0287***
	(0.004)	(0.004)	(0.009)	(0.004)
	Reduced			
Predicted Connection Time	-0.0100***	-0.0109***	0.00118	-0.000851
	(0.003)	(0.003)	(0.006)	(0.004)
	2SLS			
Real Time	-0.0406***	-0.0435***	0.00313	-0.00346
	(0.015)	(0.016)	(0.017)	(0.016)
	First Stage			
Predicted Connection Time	0.247***	0.250***	0.377***	0.246***
	(0.024)	(0.024)	(0.040)	(0.031)
Observations	4,532	4,467	1,565	2,868
R-squared	0.616	0.616	0.630	0.619
Kleibergen-Paap Wald F statistics	101.9	101.3	83.99	59.83
City fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES
Route Fixed Effects * Province Fixed Effects	YES	YES	YES	YES

Notes: Table 15 shows the robustness check. All regressions control for city fixed effects, year fixed effects, and the interaction between route fixed effects and province fixed effects. Standard errors are clustered by cities. Predicted connection time was computed by the authors using Arcmap. * p<0.05, ** p<0.01, *** p<0.001

Table 16: 2SLS: The Effect of Real Connection Time on Air Pollution Growth

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV
	Air pollution growth ($AOD_t - AODt - 1$)				
Real time	-0.008***	-0.074***	-0.076***	-0.073***	-0.076***
	(-5.083)	(-6.876)	(-6.941)	(-6.882)	(-6.954)
Log real GDP	-0.006*		-0.006		-0.007
	(-1.791)		(-1.370)		(-1.554)
Log population	-0.004			0.001	0.006
	(-0.858)			(0.216)	(0.899)
Constant	0.096***				
	(2.956)				
Observations	4,216	4,289	4,218	4,284	4,216
R-square	0.43				
Kleibergen-Paap Wald F statistics		93.26	92.27	94.15	93.84
City fixed effects	YES	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES	YES

Notes: Table 16 shows the OLS and second-stage results for the regressions using air pollution growth as the dependent variable. All regressions control for city and year fixed effects. Standard errors are clustered by 276 cities. Predicted connection time was computed by the authors using Arcmap. * p<0.05, ** p<0.01, *** p<0.001

Table 17: Heterogeneous Effects: By Region

	(1)	(2)	(3)	(4)
	Air Pollution Growth (ΔAOD)			
	Exclude municipality	Exclude other	North	South
	OLS			
Real Time	-0.00312*	-0.00259	0.00377	-0.00484**
	(0.002)	(0.002)	(0.003)	(0.002)
	Reduced			
Predicted Connection Time	-0.0178***	-0.0185***	-0.0260***	-0.0121***
	(0.002)	(0.002)	(0.004)	(0.002)
	2SLS			
Real Time	-0.0725***	-0.0743***	-0.0644***	-0.0507***
	(0.011)	(0.011)	(0.016)	(0.009)
	First Stage			
Predicted Connection Time	0.246***	0.249***	0.404***	0.238***
	(0.025)	(0.025)	(0.043)	(0.031)
Observations	4,257	4,197	1,471	2,694
R-squared	0.638	0.639	0.659	0.639
Kleibergen-Paap Wald F statistics	93.37	92.74	82.45	53.71
City fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES

Notes: Table 17 shows the heterogeneous effects by region. All regressions control for city and year fixed effects. Standard errors are clustered by cities. Predicted connection time was computed by the authors using Arcmap. * p<0.05, ** p<0.01, *** p<0.001

Table 18: Heterogeneous Effects: By Season

	(1)	(2)	(3)	(4)
	Air Pollution Growth (ΔAOD)			
	Winter	Spring	Summer	Autumn
	OLS			
Real Time	-0.00802***	0.00663**	0.00208	-0.0140***
	(0.002)	(0.003)	(0.002)	(0.003)
	Reduced			
Predicted Connection Time	-0.0119***	-0.0464***	-0.0197***	-0.000577
	(0.002)	(0.004)	(0.003)	(0.002)
	2SLS			
Real Time	-0.0485***	-0.259***	-0.0961***	-0.00258
	(0.010)	(0.041)	(0.017)	(0.007)
	First Stage			
Predicted Connection Time	0.245***	0.243***	0.267***	0.281***
	(0.025)	(0.027)	(0.028)	(0.028)
Observations	4,289	4,289	4,289	4,289
R-squared	0.638	0.402	0.410	0.419
Kleibergen-Paap Wald F statistics	93.26	70.03	79.93	88.43
City fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES

Notes: Table 18 shows the heterogeneous effects by season. All regressions control for city and year fixed effects. Standard errors are clustered by cities. Predicted connection time was computed by the authors using Arcmap. * p<0.05, ** p<0.01, *** p<0.001

Table 19: Robustness to Fixed Effects

	(1)	(2)	(3)	(4)
	Air Pollution Growth (ΔAOD)			
	Exclude municipality	Exclude other	North	South
	OLS			
Real Time	-0.00312*	-0.00259	0.00377	-0.00484**
	(0.002)	(0.002)	(0.003)	(0.002)
	Reduced			
Predicted Connection Time	-0.0178***	-0.0185***	-0.0260***	-0.0121***
	(0.002)	(0.002)	(0.004)	(0.002)
	2SLS			
Real Time	-0.0725***	-0.0743***	-0.0644***	-0.0507***
	(0.011)	(0.011)	(0.016)	(0.009)
	First Stage			
Predicted Connection Time	0.247***	0.250***	0.377***	0.246***
	(0.024)	(0.024)	(0.040)	(0.031)
Observations	4,532	4,467	1,565	2,868
R-squared	0.616	0.616	0.630	0.619
Kleibergen-Paap Wald F statistics	93.37	92.74	82.45	53.71
City fixed effects	YES	YES	YES	YES
Year fixed effects	YES	YES	YES	YES
Route fixed effects * Province fixed effects	YES	YES	YES	YES

Notes: Table 19 shows the robustness check using new fixed effects. All regressions control for city fixed effects, year fixed effects, and the interaction between route fixed effects and province fixed effects. Standard errors are clustered by cities. Predicted connection time was computed by the authors using Arcmap. * p<0.05, ** p<0.01, *** p<0.001

References

J. M. Alix-Garcia, K. R. Sims, and P. Yañez-Pagans. Only one tree from each seed? environmental effectiveness and poverty alleviation in mexico's payments for ecosystem services program. *American Economic Journal: Economic Policy*, 7(4):1–40, 2015.

D. Almond, Y. Chen, M. Greenstone, and H. Li. Winter heating or clean air? unintended impacts of china's huai river policy. *American Economic Review*, 99(2):184–90, 2009.

A. Banerjee, E. Duflo, and N. Qian. On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, page 102442, 2020.

N. Baum-Snow, L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang. Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448, 2017.

W. Blundell, G. Gowrisankaran, and A. Langer. Escalation of Scrutiny: The Gains from Dynamic Enforcement of Environmental Regulations. *American Economic Review*, 110(8):2558–2585, 2020.

R. Chen, H. Kan, B. Chen, W. Huang, Z. Bai, G. Song, and G. Pan. Association of particulate air pollution with daily mortality: the china air pollution and health effects study. *American journal of epidemiology*, 175(11):1173–1181, 2012.

O. Deschenes, H. Wang, S. Wang, and P. Zhang. The effect of air pollution on body weight and obesity: Evidence from china. *Journal of Development Economics*, page 102461, 2020.

M. Fan and G. He. The impact of clean water on infant mortality: Evidence from china. Available from `DOI:10.22004/ag.econ.291260`, 7 2020.

Y. Gao and J. Zheng. Clearing the air through pipes? an evaluation of the air pollution reduction effect of china's natural gas pipeline projects. Technical report, SSRN Working Paper No. 3475987, 2020. Available at: `doi:https://ssrn.com/abstract=3475987`.

M. Greenstone and R. Hanna. Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review*, 104(10):3038–72, 2014.

G. He, T. Liu, and M. Zhou. Straw burning, pm2. 5, and death: Evidence from china. *Journal of Development Economics*, page 102468, 2020a.

G. He, Y. Xie, and B. Zhang. Expressways, gdp, and the environment: The case of china. *Journal of Development Economics*, page 102485, 2020b.

M. Kampa and E. Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2): 362–367, 2008.

H. Kan, R. Chen, and S. Tong. Ambient air pollution, climate change, and population health in china. *Environment international*, 42:10–19, 2012.

L. Kong, J. Xin, W. Zhang, and Y. Wang. The empirical correlations between pm2. 5, pm10 and aod in the beijing metropolitan region and the pm2. 5, pm10 distributions retrieved by modis. *Environmental pollution*, 216:350–360, 2016.

P. Li, Y. Lu, and J. Wang. The effects of fuel standards on air pollution: Evidence from china. *Journal of Development Economics*, page 102488, 2020.

K. A. Miller, D. S. Siscovick, L. Sheppard, K. Shepherd, J. H. Sullivan, G. L. Anderson, and J. D. Kaufman. Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine*, 356(5):447–458, 2007.

E. Moretti and M. Neidell. Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles. *Journal of Human Resources*, 46(1):154–175, 2011.

M. J. Neidell. Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of health economics*, 23(6):1209–1236, 2004.

W. Nordhaus. Evolution of modeling of the economics of global warming: Changes in the dice model, 1992–2017. *Climatic change*, 148(4):623–640, 2018.

C. A. Pope Iii, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, 287 (9):1132–1141, 2002.

Y. Qin and K. Oduyemi. Atmospheric aerosol source identification and estimates of source contributions to air pollution in dundee, uk. *Atmospheric Environment*, 37(13):1799–1809, 2003.

N. Sastry. Forest fires, air pollution, and mortality in southeast asia. *Demography*, 39(1):1–23, 2002.

S. Sequeira, N. Nunn, and N. Qian. Immigrants and the making of america. *Review of Economic Studies*, 2019.

T. L. Sheldon and C. Sankaran. The impact of indonesian forest fires on singaporean pollution and health. *American Economic Review*, 107(5):526–29, 2017.

M. L. Sorey. *Invisible CO2 gas killing trees at Mammoth Mountain, California.* US Department of the Interior, US Geological Survey, 2000.

S. Tanaka. Environmental regulations on air pollution in China and their impact on infant mortality. *Journal of Health Economics*, 42:90–103, 2015.

S. Xu and H. A. Klaiber. The impact of new natural gas pipelines on emissions and fuel consumption in china. *Resource and Energy Economics*, 55:49–62, 2019.

J. Zhang and K. R. Smith. Household air pollution from coal and biomass fuels in china: measurements, health impacts, and interventions. *Environmental health perspectives*, 115(6):848–855, 2007.