
Air Pollution and Housing Prices: Evidence from Changes in Wind Direction at Thermal Power Plants

Shouwei Qi ^a, Yating Chen ^a, Xiumei Yu ^{a,*}

Abstract

This paper investigates the impact of air pollution on housing prices in China based on the most comprehensive second-hand housing transaction microdata and high-resolution air pollution data. To avoid the estimation biases caused by migration costs, endogeneity problem and influences of the government's price control policy on new houses, we match each house to its nearest thermal power plant and use the price of second-hand housing transaction. After controlling for the plant-year-month fixed effects and other controlling variables, we employ instrument variables based on wind direction. Results suggest that compared to upwind of the thermal power plants, PM2.5 concentrations is higher and housing price is lower downwind of the plants. Moreover, a 1% increase in PM2.5 concentrations leads to a 2.07% reduction in housing prices. Further analysis indicates that the negative impact of air pollution on housing prices is smaller in areas with better public service facilities. Our findings not only provide convincing empirical evidence on the cost of air pollution, but also indicates that in order to reduce the welfare loss caused by air pollution, in addition to reducing the pollution level through environmental regulation policies, the government can also reduce the negative impact of air pollution by improving public service facilities.

JEL Classification: Q51; Q53; H41

Keywords: Housing prices; air pollution; wind direction; thermal power plants; public service facilities

^a School of Public Finance and Taxation, Zhongnan University of Economics and Law, Wuhan, 430073, China

* Corresponding author.

E-mail addresses: shouwei.qi@gmail.com (S. Qi), yating.chen27@gmail.com (Y. Chen), xiumeiy@zuel.edu.cn (X. Yu).

I. Introduction

Air pollution is recognized as a widespread negative externality of economic activity. The literature has documented a wide array of negative impacts of air pollution, such as harms to physical health (Chay and Greenstone, 2003; Chen et al., 2013; Ebenstein et al., 2017; Chen et al., 2018; Deryugina et al., 2019) and mental health (Chen et al., 2018; Heyes and Zhu, 2019), impeded cognitive performance (Ebenstein et al., 2016; Zhang et al., 2018), decreased productivity (Zivin and Neidell, 2012; He et al., 2019; Fu et al., 2021) and reduced labor supply (Hanna and Oliva, 2015). Theoretically, we can correct such externalities and move closer to the social optimum level of pollution through several methods, including price mechanism (tax) and quantity mechanism (regulation and permits). However, in order to calculate the social optimum level of air pollution, we need to know its marginal damage, which is hard because there is no direct market to recover the willingness-to-pay. This paper aims to provide credible estimation of the marginal willingness to pay (MWTP) for clean air in China based on the most comprehensive second-hand housing transaction microdata and high-resolution air pollution data, using instrument variables developed from wind direction at thermal power plants.

Although numerous studies have estimated the implicit price of air quality in developed countries, similar studies in developing countries are limited due to data limitation (Yusuf and Resosudarmo, 2019; Freeman et al., 2019). The few existing literature based on data from developing countries like China find that households' willingness to pay for clean air is relatively low. For example, based on city-level data in China, Zhang et al. (2022) estimates that 1 unit increase in PM10 causes only 0.24% reduction in housing prices^①. The estimated low MWTP for environmental quality in developing countries is inconsistent to findings of a heavy economic and health burden generated by severe pollution, which becomes a central puzzle at the intersection of environmental and development economics (Greenstone and Jack, 2015). Therefore, a reliable estimation of the implicit price of air quality in developing countries can not only fill the gap of literature, but also help to understand the paradox between households' low valuation of air quality and high economic costs of air pollution. In addition, China as the largest

^① Based on data from the United States, Bayer et al. (2019) finds that the elasticity of housing prices with respect to PM10 is -0.63, Grainger (2012) estimates that 1 unit increase in PM10 causes 1.2% to 2% reduction in housing prices.

development countries with severe environmental concerns, has declared a “war on pollution” and undertook unprecedented regulatory changes on multiple fronts (Auffhammer et al., 2021; Greenstone et al., 2021; Karplus et al., 2021). Credible estimations of the value of environmental quality can provide empirical evidence on the benefits of environmental regulation, which is important for evaluation and future improvements of environmental policies.

Our analysis is based on the hedonic price framework proposed by Rosen (1974), which is commonly used to measure the implicit price of air quality. The intuition for the hedonic theory is that the utility derived from consuming a differentiated product, take housing as an example, is dependent on the individual characteristics of the commodity, such as air quality. Therefore, the hedonic theory predicts that air pollution has a negative impact on housing prices, with all other characteristics held constant. However, it has been shown that hedonic price studies give various estimated results on the willingness to pay for clean air, with many instances of negligible or even negative estimations.

Several econometric problems occur when implementing the hedonic price framework, especially in developing countries like China. The first concern is that households cannot move freely across places as the hedonic model assumes, especially in developing countries. The hedonic model assumes that households are free to mobile among places. As a result, people will flow from high-pollution areas to low-pollution areas, causing housing prices in high-pollution areas to fall and housing prices in low-pollution areas to rise. However, migration can be costly because moving to a new city entails both physical costs and psychological costs of leaving behind one’s family and cultural roots (Bayer et al., 2009). China’s rigid household registration system, which creates barriers for migrants to access equal opportunities as the native including public education for the kids, medical care service and qualification to purchase housing properties in some cities, making the migration cost even higher. Therefore, the hedonic price model based on city-level data is likely to draw a conclusion that housing prices are not so sensitive to local pollution. Conventional hedonic model in Freeman et al. (2019) finds an insignificant impacts of air pollution on housing prices in China, even after dealing with the endogeneity problem by instrument variables. Bayer et al. (2009) and Freeman et al. (2019) develop a residential sorting model to overcome the problem of mobility costs, raising questions whether conventional hedonic price model can be used to value environmental amenities in the

existence of migration costs.

The second concern is that the estimation can be confounded by unobservable joint determinants of air pollution and housing prices such as economic activity, which tends to result in underestimating the implicit price of air quality. In China, economic development, employment opportunities, and provision of public service facilities tend to be centralized in the same areas as polluting industries, leading to a more serious omitted variable problem (Freeman et al., 2019).

The third concern when applying the hedonic price model in China based on public available data provided by the government is that it only provides prices of new houses, which are subject to the government's price control policy. In order to prevent housing prices from rising too fast, the Chinese government has taken a series of measures to control housing prices, especially those of new houses. Real estate developers are required to provide record prices to the government when they apply for licenses to sell new houses. The transaction price must float within a certain range of the record price, making it varies little among houses sold on different dates in the same community. Therefore, the price of new houses cannot reflect market supply and demand timely.

All the three concerns result in underestimate of the MWTP for clean air, which to some extent can explain the puzzle raised by Greenstone and Jack (2015). To avoid the estimation biases caused by migration costs and the government's price control policy, instead of using city-level housing price of new houses, we use second-hand housing transaction microdata and treat houses as observations. We match each house to its nearest thermal power plant and drop those houses whose nearest power plants is beyond 6km. After controlling for the plant-year-month fixed effects, our estimations are identified from the differences of air pollution around the same thermal power plant. While migration costs across cities is considerable, moving within a city, especially around a plant is easy, making the migration cost negligible. Moreover, unlike sales of new houses, which are concentrated in certain periods, second-hand housing transactions are negotiated by buyers and sellers themselves and are carried out dispersedly. This not only avoids the impact of new houses' price controls, but also allows us to estimate the impact of pollution on a longer time scale.

Regarding the issue of endogeneity, since our estimates are based on the heterogeneity of air pollution around the same thermal power plant, the instrumental variables at the city or county level, such as the status in a certain environmental regulation policy adopted by Chay and

Greenstone (2005) and Grainger (2012), will no longer be applicable. We use the minimum angle between the local prevailing monthly wind direction and the line connecting these two locations (i.e., the housing and its nearest thermal power plant), as an instrument variable for air pollution. The intuition is that air pollution is lower upwind of the thermal power plant and higher downwind of the thermal power plant. After controlling for the plant-year-month fixed effects, it's reasonable to assume that the unpredictable and random changes in wind direction are unrelated to changes in housing prices except through their influence on air pollution.

Specially, our analysis is based on the most comprehensive second-hand housing transaction microdata in 88 cities from one of the largest housing transaction platforms in China. We match each house to its nearest thermal power plant, the list of which is provided by the government, and drop those houses to which the nearest thermal power plant is beyond 6km. After controlling the plant-year-month fixed effects and other control variables, we use the minimum angle between the local prevailing monthly wind direction and the line connecting the house and the plant as an instrument for PM_{2.5} around the house. Our results show that going from upwind to downwind of the plants, PM_{2.5} get higher and housing prices get lower. Moreover, the elasticity of housing prices with respect to PM_{2.5} is as large as -2.07, which indicates that one percentage increase in PM_{2.5} will result in 2.07 percentage decrease in housing prices. Our results are robust to different instruments and other model specifications. Further analysis indicates that the negative impact of air pollution on housing prices is smaller in areas with better public service facilities.

Our paper contributes to the literature at least in the following two aspects. Firstly, it provides substantial new evidence of the implicit price of air quality in developing countries like China. Taking advantage of our data, our identification strategy allows us to avoid estimation biases caused by migration costs, endogeneity problem, and the government's price control policy in the framework of conventional hedonic price model. The results not only demonstrate that the conventional hedonic price model can be successfully applied to value environmental amenities even in the presence of the three notable possible biases, but also provide credible empirical evidence on the economic costs of air pollution, which is important to both policy makers and economists. Secondly, the heterogeneity analysis by public service facilities contribute to the literature of public adaptation. The current literature mainly studies adaptive behaviors to

pollution from the perspective of households, such as the purchase of air purifiers (Ito and Zhang, 2020) and facemasks (Zhang and Mu, 2018), reducing outdoor activities (Zivin and Neidell, 2009; Chang et al., 2019), increasing pharmaceutical purchases (Deschenes et al., 2017) and migration (Qin and Zhu, 2018; Chen et al., 2022). Our results show that the government can also reduce the marginal costs of air pollution by improving public service facilities, which implies that in order to reduce the welfare loss caused by air pollution, in addition to directly reducing the pollution level through environmental regulation policies, the government can also reduce the negative impact of air pollution by providing better public service facilities.

II. Hedonic price theory and empirical identification challenges

II. A Hedonic price theory

A house can be thought of as a differentiated product with many individual characteristics, including housing physical attributes (e.g., housing space, number of bedrooms and parlors), the provision of public services (e.g., local transportation, educational, environmental and medical resources), and local amenities (e.g., air quality). All of these characteristics will affect the utility that households derive from consuming the house. If the mobility cost is zero, the change in utility could cause households to migrate across regions, thereby affecting the demand for local housing, and finally influencing the equilibrium price of the housing market. However, since these characteristics are not priced individually, households would not know the price they pay for each one.

The hedonic price model first proposed by Rosen (1974) has been widely used in the literature to reveal the implicit price of each characteristic of houses. In the hedonic framework, the price of each house can be expressed as a function of its characteristics, which can be written as:

$$P_h = P(q_1, q_2, \dots, q_n) \quad (1)$$

The partial derivative of $P(\cdot)$ with respect to the n th characteristic, $\partial P_h / \partial q_n$ is referred to as the implicit price of this characteristic.

Air quality around the neighborhood can be seen as one of the characteristics of a house. Since air pollution reduces households' utility through its negative impacts on life and production, it decreases the demand for houses, which in turn lowers local housing prices. The hedonic price model thus predicts that air pollution will have a negative impact on housing prices. Empirically,

equation (1) is estimated by using housing prices as dependent variables and all housing-related characteristics as independent variables. The absolute value of the air pollution coefficient reflects the marginal implicit price of air quality, which equals to people's MWTP for clean air.

II.B Empirical identification challenges

II.B.1 Estimation bias caused by migration costs

In theoretical circumstance on hedonic estimations, households are assumed to possess complete information and to be able to move freely. If moving is costless, the implicit price of air quality can be unbiased revealed by the hedonic price model. Whereby, when migration costs are high, changes in air pollution and housing prices in different areas will no longer reflect the authentic implicit price of air quality. Because the benefit that households gaining improved air quality from moving must compensate their migration costs. And existing empirical studies have neglected the migration costs problem due to their limitation of city-level data, resulting in a biased estimation.

To deal with the migration costs problem, we limit the moving distance within a certain range, employing houses adjacent thermal power plants to estimate the implicit price of air quality. The migration costs associated with intercity moving are substantial, while intracity are relatively minimal. Particularly, migration costs can be negligible in the vicinity of thermal power plants. Although the small scope of geographic space can satisfy the assumption that migration costs are negligible, it also brings unidentifiable problems. For instance, it may raise the possibility of non-existent changes in air quality, which is an important reason why previous relevant literature was limited to the city-level. However, this study is conducted about thermal power plants, which are a major source of urban air pollution. The uneven spatial distribution of thermal power plant's pollution emissions can provide a quasi-natural experiment for identifying the causal effects of air quality changes.

II.B.2 Estimation bias caused by endogeneity problems

Endogeneity is a common problem in air pollution research. For instance, unobservable socio-economic confounding factors, and measurement errors that arise from the discrepancies between the observed values at monitoring stations and the actual local air pollution levels (Greenstone et al., 2021). Currie et al. (2015) and Banzhaf (2021) have shown that in the neighborhood nearby thermal power plants, air pollution is higher and housing price is lower. To

estimate the impacts of air pollution on housing price, the distance to the nearest thermal power plant is an optimal instrument variable, but it's very likely that it does not satisfy the exclusion restriction because distance to the nearest thermal power plant can affect housing price by other channels in addition to air pollution, such as noise and water pollution. Hence, in order to address the endogeneity of air pollution, we use wind direction at thermal power plants to develop instrumental variables.

Many studies have shown that wind direction is an excellent instrumental variable for air pollution (Deryugina et al., 2019; Freeman et al., 2019). First, there is numbers of variation in wind direction, driven by large-scale weather systems and numerous unpredictable local influences, which can be considered random. Second, the variation in wind direction is an important determinant of air pollution level in the vicinity of thermal power plants. Third, wind direction at thermal power plants only affects housing prices via the distribution of air pollution, rather than other ways.

Since wind direction directly affects the distribution of air pollution in the vicinity of thermal power plants, we use air pollution from different geographical angle at plants for identification. As depicted in Figure 1, geographical angle (for short Angle) presents the minimum angle between local prevailing wind direction and the line connecting the house to its nearest thermal power plant, and its value ranges from 0° to 180° . The wind sign is indicative of local prevailing wind direction, and thin arrow-line is the extension line of the wind sign. The thick solid line indicates the line connecting the house to its nearest thermal power plant. Angle is obtained by taking the minimum angle between the thin arrow-line and the thick solid line. To satisfy the assumption that migration costs is negligible in the vicinity of the thermal power plant, we have imposed a constraint of 6-kilometers^① on the maximum distance between the house and its nearest thermal power plant. The distance calculation is performed basing on latitude and longitude information of thermal power plants and houses.

^① We take the thermal power plant as the geographical center and the radius of a certain range to form a housing submarket. We limit the distance between houses and its nearest power plant to 6km. The reason for choosing 6km is that, first, after calculating the distance between houses and all thermal power plants in the city, we get that the median of the shortest distance between houses and their nearest thermal power plant is 6.433 km, and the value of 6 km is obtained after rounding. Second, our selection is based on the perspective of government public service facilities. For instance, in China's most cities, the starting price of subway covers a rideable distance of 6-kilometers. It means that within 6km, households should not pay for travel beyond the basic price.

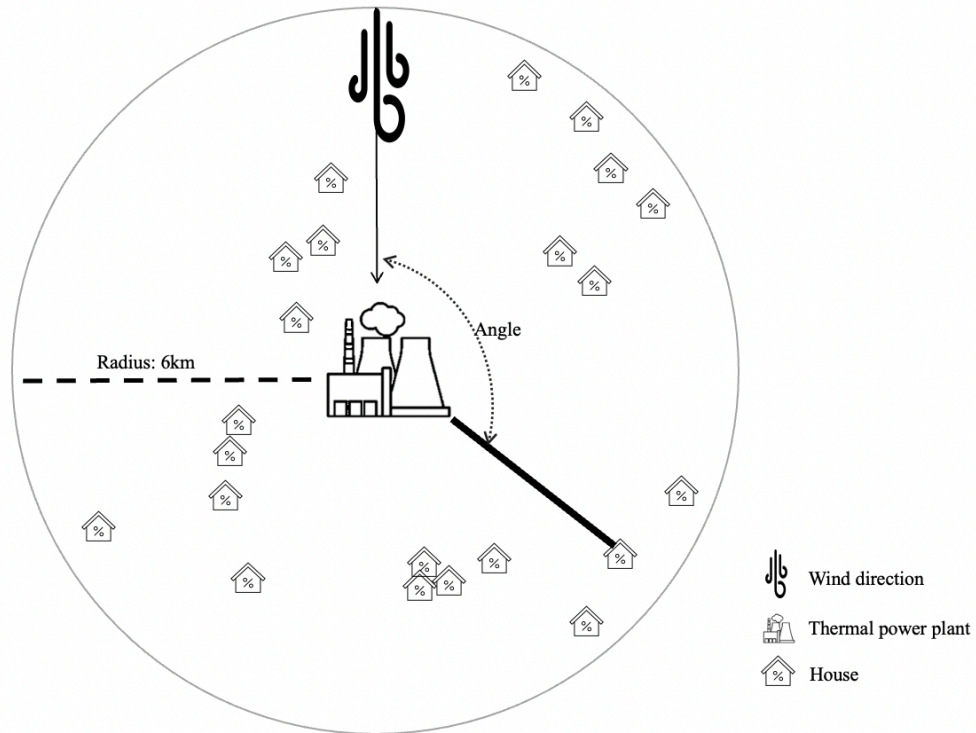


Figure 1: Schematic diagram of Angle

Designs based on quasi-natural experiments can better manage the endogeneity of the causal relationship between air pollution and housing prices and improve the accuracy of estimates. According to the relevant empirical research on the using of wind direction (Freeman., 2019; Deryugina et al., 2019; Han and Zhao, 2021), Table 1 develops five instrumental variables to address the endogeneity problem of air pollution (See the specific construction of instrumental variables in the appendix A).

Table 1: Instrumental variables developed from wind direction

Instrumental variable	Explanation
Distinct Angle IV	The minimum angle between monthly wind direction and the line connecting the house and its nearest thermal power plant (Monthly Angle for short).
45°interval Angle IV	Divide Monthly Angle into four 45 degrees intervals, representing the geographic location of the housing relative to its nearest thermal power plant.
60°interval Angle IV	Divide Monthly Angle into three 60 degrees intervals, representing the geographic location of

the housing relative to its nearest thermal power plant.

45°Bins Angle IV The number of days during the transaction month in which the daily angle (i.e., the minimum angle between daily wind direction and the line connecting the house and its nearest thermal power plant, Daily Angle for short) falls within four 45 degrees intervals, respectively.

60°Bins Angle IV The number of days during the transaction month in which Daily Angle falls within three 60 degrees intervals, respectively.

II.B.3 Estimation bias caused by government's price control policy

Previous research had primarily relied on first-hand housing data when estimating the implicit price of air quality in China, carrying an estimation bias. A major reason for this bias is the Chinese the government's price control policy on first-hand houses, resulting in first-hand housing transactions will fail to fully reflect market information.

Unlike sales of new houses, the prices of second-hand housing are determined through a negotiation process between consumers and sellers, thus the second-hand housing market is closer to a perfectly competitive market. In other words, under the situation that price regulation on first-hand houses is common in China, the second-hand housing market is relatively more competitive and can more accurately reveal the valuation of housing prices with respect to a certain characteristic. Not only that, unlike first-hand housing transactions that are concentrated in certain periods, second-hand housing transactions are negotiated by buyers and sellers themselves and are carried out dispersedly, which allows us to estimate the impact of air pollution on a longer time scale. Therefore, this study will use second-hand housing transactions to estimate the impact of air pollution on housing prices, basing on the hedonic price model modified by instrumental variable methods.

III. Data sources and descriptive statistics

III.A Data sources

III.A.1 Housing transaction data

Our second-hand housing data come from the "Lianjia" transaction platform, which is famous for providing comprehensive and high-quality information. In Beijing, "Lianjia" holds a

55%-60% share of the local housing intermediary market. In other cities such as Tianjin, Qingdao, Chengdu, and Nanjing, "Lianjia" is also ranking first or second in market share. The housing data from 2012 to 2020 covers a total of 88 cities across China, which across all provinces, autonomous regions, and municipalities in China, except for Guizhou, Yunnan, Gansu, Qinghai, Tibet, and Xinjiang. The spatial distribution of housing samples is represented in Figure 2, wherein Beijing, Chengdu, Shanghai, and Tianjin are owning the largest housing sample size, which is consistent with the trends observed in macroeconomic indicators.

The housing data include total transaction price, transaction unit price, total listing price, transaction date, number of bedrooms, number of parlors, living space, structure, degree of decoration, availability of elevators, orientation, and address information. We limit our analysis to the samples pertaining to general residential and commercial houses and exclude those with incomplete information. We winsorize housing prices and living space separately at the 1% level, aiming to control for the impact of outliers on the accuracy of the estimates. Following this data processing step, the resulting observations involves 1,470,815 housing transactions.

Then, we use the latitude and longitude information to match each house with its closest thermal power plant. Based on the matching results, we removed houses that were more than 6 km away from the nearest thermal power plant, resulting in the ultimate analysis observations, that is, 583,021 housing transactions.

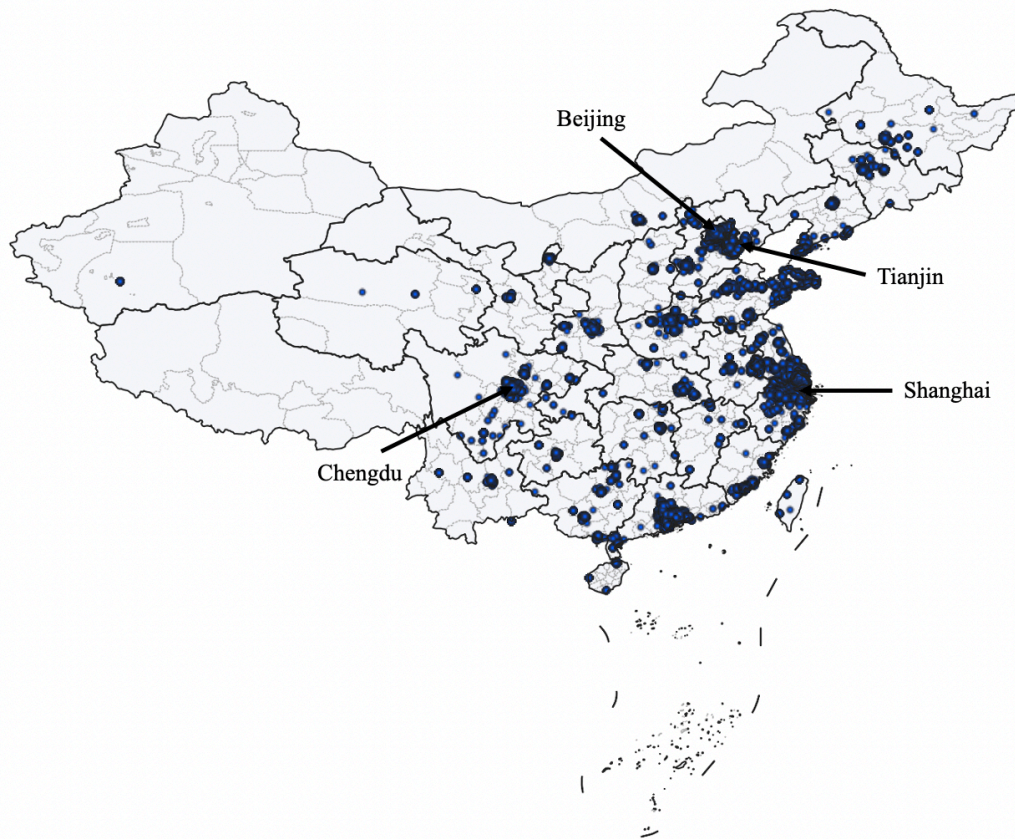


Figure 2: Spatial distribution of housing samples

Note:

(1) Figure 2 displays the top four cities with the largest housing sample size in the original data, the first one is Beijing with 422,676 samples; the second one is Chengdu with 211,378 samples; the third one is Shanghai with 136,472 samples; the fourth one is Tianjin with 106,189 samples.

(2) The solid black line represents the administrative boundary line at the provincial level, and the dotted gray line represents the administrative boundary line at the municipal level. The blue dot represents the housing geographic location. The co-location of certain dots on the map is because houses located within the same community possess latitude and longitude data that is exceptionally alike, thereby resulting in overlapping representations.

III.A.2 Thermal power plants data

The burning of fossil fuels in thermal power plants constitutes a significant source of air pollution in cities. Thermal power plants account for over 63% of China's electricity generation^①, but also contribute to 60% of the country's industrial pollution emissions (Zheng and Kahn, 2017). The list of thermal power plants used in this study comes from Ministry of Ecology and

^① According to the BP report, China's coal-fired power generation will account for 4917.7/7779.1=63.2% of China's power generation in 2020; China's coal-fired power generation accounts for 4917.7/9421.4=52.2% of the global coal-fired power generation; China's coal-fired power generation accounts for 4917.7/26823.2=18.3% of the world's total power generation. Data sources: https://www.sohu.com/a/480454569_257552.

Environment of the People's Republic of China^①.

As part of the national carbon emissions trading quota management program, Ministry of Ecology and Environment of the People's Republic of China (2020) has compiled a list of key emission units. The list encompasses power generation industries, including self-owned power plants in other sectors, that emitted over 26,000 tons of carbon dioxide equivalent in any given year spanning 2013-2019. This emission is equivalent to the comprehensive energy consumption of around 10,000 tons of standard coal. Following a rigorous screening process, 2,225 thermal power plants were found to be subject to inventory management^②. Thus, while the list was originally created for China's national carbon emissions trading market, it covers nearly all the China's major thermal power plants, for instance, the Yangcheng Power Plant in Shanxi with 7,260 MW capacity, Tuoketuo Power Plant in Inner Mongolia with 6,720 MW capacity, and Waigaoqiao Power Plant in Shanghai with 5,000 MW capacity.

After obtaining the list of thermal power plants, we retrieve their detailed addresses from the annual report based on the name of each plant. Subsequently, we use the Google Maps Platform's Application Programming Interface (API) to extract the geographic location (i.e., latitude and longitude information) of each thermal power plant from its detailed address.

III.A.3 Air pollution data

This study uses PM_{2.5} concentrations (Hereinafter referred to as PM_{2.5}) as a proxy variable for air pollution. PM_{2.5} represents airborne particulate matter with an aerodynamic equivalent diameter smaller than 2.5 μm (Dominici et al., 2014). Empirical studies have shown that PM_{2.5} have negative impacts on health (Dominici et al., 2014; Deryugina et al., 2019), impeded cognitive performance (Ebenstein et al., 2016), decreased productivity (Fu et al., 2021), etc.

To satisfy the needs of this paper to measure air pollution changes in a fine geographic space, we use PM_{2.5} data come from the Monthly Global Estimates of Fine Particulate Matter derived from satellite-based Aerosol Optical Depth (AOD) retrieval techniques by Van Donkelaar et al. (2021)^③. AOD essentially measures the amount of sunshine duration that are absorbed, reflected,

^① Sources: https://www.mee.gov.cn/xxgk/2018/xxgk/xxgk03/202012/t20201230_815546.html.

^② Thermal power plants in the list can be divided into four types of unit categories, the first is conventional coal-fired units above 300MW, the second is the 300MW level and the conventional coal-fired units below, and the third is unconventional coal-fired units such as coal-fired vermiculite and coal slurry (contains coal-fired circulating fluidized beds), the fourth is the gas unit. The classification basis comes from: <https://www.mee.gov.cn/xxgk/2018/xxgk/xxgk03/202012/W020201230736907121045.pdf>.

^③ Van Donkelaar et al. (2021) estimate monthly ground-level fine particulate matter (PM_{2.5}) for 1998-2020 by combining

and scattered by the particulates suspended in the air, and can be used to estimate fine particular matter concentrations. Gupta et al. (2006), Kumar et al. (2011) have verified the accuracy of the satellite-based data using ground-based, station data in China. They have concluded that the observed differences between the two datasets are not statistically significant, given the incorporation of geographic and year fixed effects.

The satellite-based data used in this study have several advantages compared to ground-based air pollution data. Including, ground-based data confine to only larger and mid-sized cities, but the satellite-based data provide the most comprehensive measures of PM_{2.5} across China's geography. In other words, the satellite-based data can be used to predict PM_{2.5} even in areas lacking ground-based monitoring stations. Besides, the satellite-based data exhibit a high reliability, owing to their generation process that is devoid of human interference. Third, the satellite-based data have a high grid cell resolution of 0.01°x0.01° degree, allowing for the accurate measuring of PM_{2.5} at the scale of one square kilometer.

III.A.4 Wind direction data

The wind direction data used in this study are obtained from the National Climatic Data Center (NCDC), which has been a leading source of meteorological data for global weather stations since 1942. The NCDC dataset is subject to rigorous quality control procedures, resulting in a missing rate of less than 1% and an accuracy rate approaching 100%. The wind direction data provided by NCDC is reported in degrees spanning 0°~360°, which is essential for accurately calculating of instrument variables in this study.

A selection rule is developed for selecting appropriate weather stations from which to obtain wind direction at thermal power plants. The first step is to calculate the distance between the thermal power plant and all of weather stations across China, and then only keep the weather station nearest to the thermal power plant following a sorting process. After that, we can obtain daily prevailing wind direction at the thermal power plant from its nearest weather station. Lastly, we convert wind direction from the daily level to the monthly level by using the unit vector method. Instructions for the unit vector method are explicated in the appendix B.

Aerosol Optical Depth (AOD) retrievals from the NASA MODIS, MISR, and SeaWiFS instruments with the GEOS-Chem chemical transport model, and subsequently calibrating to global ground-based observations using a Geographically Weighted Regression (GWR).

III.A.5 Public service facilities data

In the conventional house price determinant model, public service facilities are regarded as key factors. To control the potential impact of public service facilities on housing prices, we gathered relevant data from Baidu Maps' collection of points of interest (POI). POI refers to point data of some landmark buildings and geographical entities that are closely associated with human livelihood, including schools, hospitals, parks, and government agencies. POI data describes the spatial location and attribute characteristics of these geographical entities. It has a large sample size and abundant information, thus can effectively reflect various activities of the city to a certain extent. Generally, each POI record comprises four key elements, name, category, latitude and longitude, and address.

Public service data used in this study come from four categories, transportation resources (subway stations), educational resources (primary schools), environmental resources (parks), and medical care resources (general hospitals). We use the one-to-many method to calculate the straight-line distance between each housing observation and all public service samples through their latitude and longitude information. To enhance computational efficiency and reduce processing time of the distance calculation process, we perform it for each city-year combination.

III.A.6 Economic activity data

We use Visible and Infrared Imaging Suite (VIIRS) nighttime light data as a proxy variable for economic activity around the house, obtained from Earth Observation Group (EOG)^①. As the pioneer of the nocturnal remote sensing technology, Earth Observation Group had been collecting nighttime satellite imagery and produce monthly and annual global nighttime light map with high quality, whose history can trace back early as 1994. There are a substantial number of steps involved in producing VIIRS nighttime light data, such as cleaning steps to exclude background noise, solar and lunar contamination, data degraded by cloud cover, and features unrelated to electric lighting (e.g., fires, flares, volcanoes) (Elvidge et al., 2017). Empirical research has also proved that VIIRS nighttime light data has good consistency in the simulation of economic indicators, population indicators, energy consumption indicators and land use type indicators (Shi

^① <https://eogdata.mines.edu/products/vnl/>.

et al., 2014; Zheng et al., 2017). A VIIRS nighttime light image has an image resolution of approximately 0.5 km at the equator, thereby enabling the identification of variations in economic activity across small geographic areas.

III.B Descriptive statistics

Figure 3 displays the reliability of our identification from the aspect of data correlation. Figure 3a represents the spatial distribution of air pollution, which is positively related to Angle; and Figure 3b represents the negative relationship between housing prices and Angle. To avoid the regional differences in air pollution and housing prices between the locations of thermal power plants, as well as the time trend effects, we have removed the plant-year-month fixed effects^① from the air pollution in Figure 3a and housing prices in Figure 3b. In Figure 3a, the y-axis represents the residual values of air pollution after removing the plant-year-month fixed effects, while in Figure 3b, the y-axis represents the residual values of housing prices after removing the plant-year-month fixed effects. Both figures take Angle as the x-axis. As demonstrated by the results, there is an apparent increase in air pollution in the areas downwind of thermal power plants (i.e., Angle increasing), which is accompanied by an apparent decline in housing prices.

^① Fixed effects usually refer to the portion of an empirical analysis that varies due to inherent differences between sample s , which include environmental background, economic status, cultural traditions, etc. Failure to exclude these fixed effects leads to confounding effects, making it difficult to correctly estimate the relationship between variables.

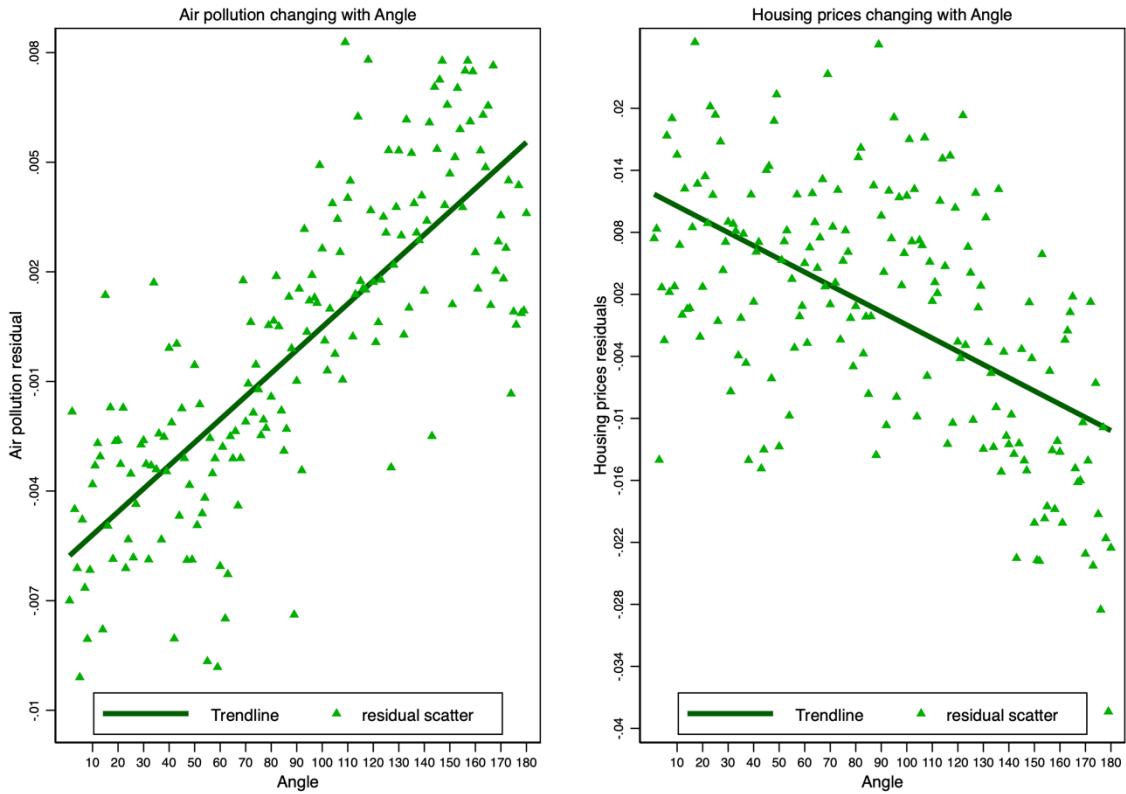


Figure 3: Trend of air pollution and housing prices as Angle change

Note: Starting from the left, the first figure is defined as Figure 3a and the second figure as Figure 3b.

Table 2 presents descriptive statistics of the key variables, all of them are at the housing-month level. The dependent variable in this study is transaction unit price of the house. The average housing price is 34,649 yuan/m² and the median is 26,234 yuan/m², which is consistent with the trend of housing prices in the macro data. The independent variable with which this study is concerned is PM_{2.5} at the housing level. The average PM_{2.5} that all housing samples reached 55.46 ug/m³, a figure much higher than the daily standard of World Health Organization (WHO) standard of 25 ug/m³, and the annual standard of 10 ug/m³. It means that houses adjacent thermal power plants are exposed to greater air pollution.

To observe the spatial distribution of the housing samples around thermal power plants, we have divided Angles into four groups, each spanning an interval of 45 degrees. As shown in Table 2, the probability ratio of Angle=[0,45), Angle=[45,90), Angle=[90,135), and Angle=[135,180) is nearly the same, as evidenced by the average values of 0.231, 0.255, 0.256, and 0.257, respectively. This suggests that the spatial distribution of housing samples does not lead to bias

in the estimated results of this study.

As for control variables, in addition to the influencing factors of housing characteristics, we further matched the POI data to houses to control the premium of public service facilities included in housing prices and matched VIIRS nighttime light data to control the impact derived from economic activity.

Table 2: Descriptive statistics of the main variables

Variable	Variable definition	N	mean	sd	min	p50	max
Dependent variable							
Price	Housing transaction unit prices, unit: yuan/m ²	583,021	34,649	25,691	5,223	26,234	113,711
Independent variable							
PM2.5	PM2.5 concentrations around the house, unit: ug/m ³	583,021	55.46	25.66	8.600	51.20	175.7
Instrumental variables							
Angle		583,021	91.92	51.34	0.0012	92.46	180.0
Angle=[0,45)		583,021	0.231	0.422	0	0	1
Angle=[45,90)		583,021	0.255	0.436	0	0	1
Angle=[90,135)		583,021	0.256	0.436	0	0	1
Angle=[135,180]		583,021	0.257	0.437	0	0	1
Control variables							
Listing-price	Housing listing prices; unit: yuan/m ²	526,504	33,412	25,723	5,171	23,980	119,048
Economic	Economic activity level around the house	583,021	9.413	5.979	0	10	26
Space	Housing living space; unit: m ²	583,021	83.36	35.77	31	77.71	219.5
Room	Number of housing bedrooms	583,021	2.114	0.817	0	2	9
Parlors	Number of housing parlors	583,021	1.308	0.580	0	1	5
Subway	Dummy variable of whether there is a subway station within 0.5km of the house	583,021	0.289	0.453	0	0	1
School	Dummy variable of whether there is a primary school within 0.5km of the house	583,021	0.550	0.497	0	1	1

Park	Dummy variable of whether there is a park within 0.5km of the house	583,021	0.463	0.499	0	0	1
Hospital	Dummy variable of whether there is a general hospital within 0.5km of the house	583,021	0.490	0.500	0	0	1
Decoration	Dummy variable of housing degree of decoration	583,021	Rough room; simple room; hardcover room Brick-wood structure; brick-concrete structure;				
Structure	Dummy variable for type of housing structure	583,021	hybrid structure; frame structure; steel-concrete structure; steel structure				
Elevator	Dummy variable for availability of elevators of the house	583,021	The value is 1 if the elevator is installed, and the value is 0 if the elevator is not installed				
Toward	Dummy variable for housing orientation	583,021	Orientation of housing in eight directions				

Note:

(1) Air pollution data come from the Monthly Global Estimates of Fine Particulate Matter derived from satellite-based Aerosol Optical Depth (AOD) retrieval techniques by Van Donkelaar et al. (2021).

(2) Housing data come from the "Lianjia" online transaction platform.

(3) Public service facilities data come from Baidu POI.

(4) Economic activity data come from VIIRS nighttime light data provided by Earth Observation Group.

IV. Econometrics specifications

IV.A OLS method

High migration costs mean that the conventional hedonic model will be unable to recover unbiased estimates of the implicit price of air quality. For this reason, this study uses the hedonic price model within a quasi-experimental design, whereby the research sample is confined to houses adjacent thermal power plants, with the aim of decreasing possible estimation biases arising from migration costs. The Ordinary Least Square (OLS) method is applied as the basic econometric approach, and fixed-effect panel models are further applied to control the effects of missing bias, yields the following equation:

$$\ln Price_{p,h,y,m} = \alpha + \beta \ln Pollt_{p,h,y,m} + \lambda^k Controls_{p,h,y,m}^k + \gamma_p * \theta_y * \delta_m + \varepsilon_{p,h,y,m} \quad (2)$$

Where h denotes individual housing, p thermal power plant, y year and m month. For

each housing sample h , $\ln Price$ represents the logarithm of housing unit price, $\ln Pollt$ represents the logarithm of PM2.5 in the geographical location of the house. The coefficient β captures the elasticity of housing prices with respect to PM2.5, which approximate interpretation suggests that for every 1% increase in PM2.5 leads to a $\beta\%$ change in housing prices.

In addition to air pollution, numerous other factors exert a substantial influence on housing prices. *Controls* represent variables that influence housing prices in housing price determinants model, which can be categorized into three groups. The first group pertains to physical characteristics of the house, such as living space, number of bedrooms, number of parlors, degree of decoration, structure, availability of elevators, and orientation. The second group encompasses public service facilities near the house, including proximity to subway stations, elementary schools, parks, and hospitals. The third group involves economic activity around the house, which estimated by a common-used indicator, that is VIIRS nighttime light data. We apply a logarithmic transformation to VIIRS nighttime light data to effectively mitigate the abrupt increase in radiation intensity within the urban core.

Fixed effects ($\gamma_p * \theta_y * \delta_m$) capture time-persistent plant characteristics that affect housing prices, and simultaneously capture year-month shocks to housing prices such as business cycle or macroeconomic effects. Since in our sample time period, all thermal power plants have been built before the first period, that is, the thermal power plant closest to a particular house will not change. Thus, adding $\gamma_p * \theta_y * \delta_m$ that including plant fixed effects can absorb most of the impact on housing prices space time-invariant and plant-specific factors. We would like to highlight that, even after including these FEs, the cross-sectional housing prices and air pollution variation in the model is large enough to identify meaningful elasticity coefficients. The error term ($\varepsilon_{p,h,y,m}$) captures time-varying, plant-specific factors that affect housing prices.

We aggregate the monthly air pollution and wind direction to the thermal power plant level because the geographic location of all houses can be well known at the plant level. Because of this, we cluster standard errors by plant-year-month in all estimation to allow for serial correlation of factors over time within plant and spatial correlation of factors within each year-month.

IV.B Instrumental variable methods

Our objective is to estimate the impact of air pollution on housing prices, while controlling for potential confounding factors, and additionally satisfy the assumption of negligible migration

costs. OLS estimates will bias the impact of air pollution on housing prices upward or above zero due to reverse causality and missing variables (Chen and Jin, 2019). Thus, we employ instrumental variable (IV) methods to model air pollution, in order to address the endogeneity problem.

IV.B.1 Reduced form specification

For an instrumental variable to be considered effective, it is necessary to satisfy the assumption that it should exhibit a robust correlation with the variable that it is being instrumented. Thus, we use instrumental variables to replace the independent variables of the original equation, directly showing the correlation between the choice of instrumental variables and the core dependent variable. The econometric specification of reduced form is expressed as following equation:

$$\ln Price_{p,h,y,m} = \alpha' + \beta' Z_{p,h,y,m} + \lambda^k Controls_{p,h,y,m}^k + \gamma_p * \theta_y * \delta_m + \varepsilon_{p,h,y,m}' \quad (3)$$

Where Z represents the five types of instrumental variables as shown in Table 1. The remaining variables are explained in the same way as Equation (2).

IV.B.2 IV-2SLS specifications

We then use the following first-stage equation of IV-2SLS methods to model PM2.5 at the housing level, $\widehat{\ln Pollt}_{p,h,y,m}$, in Equation (4):

$$\widehat{\ln Pollt}_{p,h,y,m} = \alpha'' + \beta'' Z_{p,h,y,m} + \lambda^k Controls_{p,h,y,m}^k + \gamma_p * \theta_y * \delta_m + \varepsilon_{p,h,y,m}'' \quad (4)$$

The second stage of IV-2SLS methods is carried out by substituting the modeling PM2.5 from the first stage into Equation (3). The combination of equations after substitution is well solved the endogenous problem, so it can accurately calculate the elasticity of housing prices with respect to PM2.5 changes. The resulting econometric specification can be expressed as follows:

$$\ln Price_{p,h,y,m} = \alpha''' + \beta''' \widehat{\ln Pollt}_{p,h,y,m} + \lambda^k Controls_{p,h,y,m}^k + \gamma_p * \theta_y * \delta_m + \varepsilon_{p,h,y,m}''' \quad (5)$$

Where, besides the independent variables PM2.5 ($\widehat{\ln Pollt}_{p,h,y,m}$), we also include the same control variables ($Controls_{p,h,y,m}$), fixed effects ($\gamma_p * \theta_y * \delta_m$), error term ($\varepsilon_{p,h,y,m}$) and cluster

standard errors like Equation (3).

VI. Empirical results

VI.A OLS estimated results

In Table 3, we use Equation (2) to estimate the elasticity of housing prices with respect to PM2.5, which is revealed by the coefficient of Ln(PM2.5). Column (1) yields an elasticity of 0.34 without controlling for fixed effects, while Columns (2)-(4) yield estimated elasticities of 0.05~0.61, depending on how the fixed effects are controlled. Conventional hedonic price model for estimating the implicit price of air quality relies on the assumption that households move freely across places. We indicate that when moving is costly, the variation in housing prices and air pollution across places may no longer reflect the authentic relationship. While it is possible to tackle the problem of migration costs by only using houses adjacent thermal power plants, our findings indicate that the OLS estimated results are at odds with our intuition, but align with the results presented in Freeman et al. (2019). Consequently, relying on OLS to estimate the elasticity may introduce bias into our analysis. To address these concerns and mitigate potential estimation bias, we employ instrumental variables developed from wind direction at thermal power plants to model air pollution.

Table 3: OLS estimated results of air pollution on housing prices

	(1)	(2)	(3)	(4)
Ln(PM2.5)	0.340***	0.609***	0.049***	0.080***
	(0.026)	(0.039)	(0.011)	(0.026)
Controls	Yes	Yes	Yes	Yes
Year*Month FE	No	Yes	Yes	No
Plant FE	No	No	Yes	No
Plant*Year*Month FE	No	No	No	Yes
<i>N</i>	583,021	583,021	583,021	583,021
<i>R</i> ²	0.287	0.334	0.878	0.892
adj. <i>R</i> ²	0.287	0.334	0.878	0.891

Notes: Each observation represents a house. The regression of Table 3 is based on the hedonic price model, with the natural logarithm of housing prices as the dependent variable, the natural logarithm of PM2.5 as the main independent variable. All regressions include control variables for housing characteristics, public service facilities and local economic activity. Fixed effects

are added stepwise. Standard errors are clustered at the plant-year-month level. Standard errors in parentheses, *p<0.1, **p<0.05, ***p<0.01.

VI.B IV estimated results

VI.B.1 Estimated results of the reduced form

Table 4 use Equation (3) to perform regression, showing the reduced form estimated results for the IV methods. The simplified form is to estimate the effect of endogenous variables on the causal effect directly by exploiting instrument variables, while controlling for the effects of other variables. If the results of the reduced form regression show that an instrument variable has a significant effect on the dependent variable, then this instrument variable can be considered valid and the IV-2SLS method can be used to estimate the causal effect.

The estimate results of our baseline IV method (Distinct Angle IV) are presented in Column (1), which reveals an inverse correlation between Angle and housing prices. Specifically, a one-degree increment in Angle is linked to a statistically significant 0.02% decline in housing prices at the 1% significance level. The estimated results of remaining IV methods are presented in Columns (2)-(5). Specifically, 45°interval Angle IV, 60°interval Angle IV, 45°Bins Angle IV and 60°Bins Angle IV all use the house located upwind of thermal power plants as the reference group. The findings consistently indicate that houses located in the areas downwind of thermal power plants display a notable disadvantage in prices compared to those located in the upwind. This empirical evidence corroborates the reliability of the results obtained from Figure 3.

Table 4: Estimated results of the reduced form

	(1)	(2)	(3)	(4)	(5)
	Distinct Angle	45°interval Angle	60°interval Angle	45°Bins Angle	60°Bins Angle
	IV	IV	IV	IV	IV
Angle	-0.0002*** (0.0000)				
Angle=[45,90)		-0.004 (0.004)		0.000 (0.000)	
Angle=[90,135)		-0.007 (0.005)		0.000 (0.000)	
Angle=[135,180)		-0.026***		-0.002***	

		(0.006)		(0.000)	
Angle=[60,120)			-0.003		0.000
			(0.004)		(0.000)
Angle=[120,180)			-0.020***		-0.002***
			(0.005)		(0.000)
Controls	Yes	Yes	Yes	Yes	Yes
Plant*Year*Month					
FE	Yes	Yes	Yes	Yes	Yes
<i>N</i>	583,021	583,021	583,021	583,021	583,021
<i>R</i> ²	0.892	0.892	0.892	0.892	0.892
<i>adj. R</i> ²	0.891	0.891	0.891	0.891	0.891

Notes: Each observation represents a house. Table 4 shows the estimated results for the reduced form of IV methods. All regressions take the natural logarithm of housing prices as the dependent variable, and the natural logarithm of PM2.5 as the independent variable. All regressions include control variables for housing characteristics, public service facilities and local economic activity. Fixed effects controlled at plant-year-month level. Standard errors are clustered at the plant-year-month level. Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

VI.B.2 Estimated results of the first stage

In Table 5, regressions are performed using Equation (4), resulting in estimated results for the first stage of IV methods. Column (1) employs Angle as the instrumental variable for PM2.5 (Distinct Angle IV), and the regression results reveal a statistically significant positive relationship between Angle and PM2.5. These results are in line with our expectations, as residences located closer to the areas downwind of thermal power plants are likely to be more exposed to PM2.5, as well as exposure levels increasing gradually following Angle.

In Columns (2)-(5), we adjust the instrumental variables for PM2.5. In Column (2), we replaced the baseline IV with 45°interval Angle IV. The reference group comprises all housing samples falling within Angle=[0,45), while all others are considered treatment groups. In Column (3), we use 60°interval Angle IV to perform analysis. By adjusting the size of treatment groups, we can estimate the robustness of the first-stage estimation. Both Columns (2)-(3) reveal a positive correlation between instrumental variables and PM2.5. In Column (4), we use the number of days of wind blowing from the upwind areas as the baseline group to estimate the impact of

45°Bins Angle IV on PM2.5; whereas Column (5) estimates the impact of 60°Bins Angle IV on PM2.5. As demonstrated in Table 5, all our instrumental variables effectively address endogeneity concerns present in conventional hedonic estimations.

Table 5: Estimated results of the first stage of IV-2SLS methods

	(1)	(2)	(3)	(4)	(5)
	Distinct	45°interval Angle	60°interval Angle	45°Bins Angle	60°Bins Angle
	Angle IV	IV	IV	IV	IV
Angle	0.0001***				
	(0.0000)				
Angle=[45,90)		0.002		0.000	
		(0.001)		(0.000)	
Angle=[90,135)		0.009***		0.000***	
		(0.002)		(0.000)	
Angle=[135,180)		0.011***		0.001***	
		(0.002)		(0.000)	
Angle=[60,120)			0.006***		0.000
			(0.001)		(0.000)
Angle=[120,180)			0.011***		0.001***
			(0.002)		(0.000)
Controls	Yes	Yes	Yes	Yes	Yes
Plant*Year*Month FE	Yes	Yes	Yes	Yes	Yes
<i>N</i>	583,021	583,021	583,021	583,021	583,021
<i>R</i> ²	0.988	0.988	0.988	0.988	0.988
<i>adj. R</i> ²	0.987	0.987	0.987	0.987	0.987

Notes: Each observation represents a house. Table 5 shows the first stage estimated results of IV-2SLS methods. All regressions take the natural logarithm of PM2.5 as the dependent variable, adding control variables and plant-year-month fixed effects. Standard errors are clustered at the plant-year-month level. Standard errors in parentheses, *p<0.1, **p<0.05, ***p<0.01.

VI.B.3 Estimated results of the second stage

In contrast to estimation from the conventional hedonic model, employing a quasi-natural experimental design enables more effective control of endogeneity in the causal linkage between

air pollution and housing prices, resulting in an improved estimation accuracy. In this section, we employ the IV-2SLS methods for estimation. Table 6 shows the estimated elasticities of housing prices with respect to PM2.5, which are derived from Equation (5).

In Column (1), we employ the Distinct Angle IV method to conduct regression analysis, and our estimated results reveal that a 1% increase in PM2.5 is correlated with a 2.07% reduction in housing prices. These results underscore the vital impact of air pollution on the housing market. When employing instrumental variables methods, it is inevitable to estimate the potential presence of weak instruments. The Kleibergen-Paap Wald rk F statistic can be viewed as the method used to test the validity of instrumental variables in the second stage of the IV-2SLS method. It provides a robust method to test the consistency and validity of IV estimates and can help researchers more accurately estimate causal effects. As shown in instrumental variable test in Table 6, the Kleibergen-Paap Wald rk F statistic of Distinct Angle IV exceeds the critical value of 16.38 established by Stock and Yogo (2005) with a 10% margin of error, indicating that our baseline IV successfully rejects the hypothesis of weak instrumental variable.

Bayer et al. (2009), Freeman et al. (2019) have also observed the migration costs problem and have proposed the residential sorting model as an alternative to the hedonic price model, which does not adequately account for the costs associated with mobility. Our estimated elasticity (2.07%) is slightly larger than that 0.37% estimated by Bayer et al. (2009), and 0.98% estimated by Freeman et al. (2019). Migration costs engender distortions in the implicit price of air quality, and the underestimation observed in previous studies has two origins. On the one side, they exclusively addressed the influence of the household registration system in relation to migration costs. However, aside from institutional barriers, physical costs and psychological costs also contribute to the aggregate migration costs. On the other side, all of them employed city-level data, thereby either limited to controlling for higher-level fixed effects than the city-level or failing to control for such effects.

Results of the second stage of the other IV methods are displayed in Columns (2)-(5) of Table 6, corresponding to 45°interval Angle IV, 60°interval Angle IV, 45°Bins Angle IV and 60°Bins Angle IV. Our findings reveal that the elasticity of housing prices with respect to PM2.5 consistently displays a negative association, with estimates ranging from -1.82% to -2.29%, and all coefficients are statistically significant at the 1% level. After considering the migration cost,

the hedonic estimation obtained by the IV method will be higher than those without considering the migration cost in the literature. Based on these results, we can draw the conclusion that the main specification used in this paper performs well, and that our findings remain robust to alternative econometric specifications employed in the second stage.

Table 6: Estimated results of the second stage of IV-2SLS methods

	(1)	(2)	(3)	(4)	(5)
	Distinct Angle	45°interval Angle	60°interval Angle	45°Bins Angle	60°Bins Angle
	IV	IV	IV	IV	IV
Ln(PM2.5)	-2.072*** (0.593)	-1.823*** (0.552)	-1.863*** (0.577)	-2.294*** (0.611)	-2.278*** (0.597)
Economic	-0.014 (0.018)	-0.012 (0.017)	-0.013 (0.018)	-0.015 (0.018)	-0.015 (0.018)
Space	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Room	0.023*** (0.002)	0.023*** (0.001)	0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)
Parlor	0.046*** (0.002)	0.046*** (0.002)	0.046*** (0.002)	0.046*** (0.002)	0.046*** (0.002)
Elevator	0.114*** (0.003)	0.114*** (0.003)	0.114*** (0.003)	0.114*** (0.004)	0.114*** (0.004)
Subway	0.042*** (0.003)	0.042*** (0.003)	0.042*** (0.003)	0.042*** (0.003)	0.042*** (0.003)
School	0.020*** (0.002)	0.019*** (0.002)	0.019*** (0.002)	0.020*** (0.003)	0.020*** (0.003)
Park	0.031*** (0.003)	0.030*** (0.003)	0.030*** (0.003)	0.031*** (0.003)	0.031*** (0.003)
Hospital	0.023*** (0.003)	0.022*** (0.003)	0.022*** (0.003)	0.023*** (0.003)	0.023*** (0.003)
Controls	Yes	Yes	Yes	Yes	Yes

Plant*Year*Month FE	Yes	Yes	Yes	Yes	Yes
Instrumental variable test					
Prob > F	0.000	0.000	0.000	0.000	0.000
Kleibergen-Paap Wald rk F statistic	35.14	12.09	17.81	13.87	21.18
Stock-Yogo weak ID test critical values					
10% maximal IV size	16.38	22.30	19.93	22.30	19.93
15% maximal IV size	8.96	12.83	11.59	12.83	11.59
20% maximal IV size	6.66	9.54	8.75	9.54	8.75
25% maximal IV size	5.53	7.80	7.25	7.80	7.25

Notes: Each observation represents a house. Table 6 shows the second stage estimated results of IV-2SLS methods. All regressions take the natural logarithm of housing prices as the dependent variable and the natural logarithm of PM2.5 as the independent variable, adding control variables and plant-year-month fixed effects. Standard errors are clustered at the plant-year-month level. Standard errors in parentheses, *p<0.1, **p<0.05, ***p<0.01.

VLC MWTP and welfare implication

We can reveal the marginal willingness to pay for clean air following the estimated elasticity of housing prices with respect to PM2.5. The method presented in this study exhibits distinctive advantages. Previous literature focused on the city-level failed to address the issue of high migration costs between cities during estimation, resulting in an underestimation of the marginal willingness to pay for clean air. Instead, our calculation is based on the elasticity that has satisfied the assumption that migration cost is negligible, thereby providing a more accurate representation of the MWTP.

A popular assumption is that households' preferences are homogeneous and linear with respect to air quality, so that the MWTP for clean air is constant (Freeman, 1974; Chay and Greenstone, 2005). In this case, calculating the MWTP is simple. In our data, the average price of second-hand housing is 34,649 yuan/m², which means that the change in housing prices brought about by a 1% change in air pollution is 717.23 yuan (34,649 yuan*2.07%).

If it is extended to the whole country, China's housing market value is 62.6 trillion dollars (418 trillion yuan) until 2020, which is 4.1 times of GDP^①. Therefore, the improvement of per

^① <http://finance.sina.com.cn/zl/china/2021-10-28/zl-iktzscyy2178849.shtml>.

percentage of air pollution can bring about an increase in economic value of 8.65 trillion yuan (418 trillion yuan*2.07%). Although credible estimates of the costs of air pollution regulation are not available, our results indicate that the authentic implicit price of air quality improvement is substantially higher than has been previously recognized, and the welfare effects of environmental quality improvements are substantial in developing countries like China. Also, this may only be the lower bound of the social benefits brought about by air quality improvements. As economic growth is accompanied by the accumulation of wealth, an increase in income, and conceptual innovation, households are increasingly inclined to attribute a higher price to clean air.

We have demonstrated that our quasi-experimental methods can yield accurate parameter estimates for IV hedonic specifications. In a broader sense, this study has an implication for forthcoming research. Our findings indicate that the hedonic price model remains applicable in the presence of migration costs, as demonstrated by our causal identification.

VI.D Robustness checks

In Column (1) of Table 7, we employ a semi-log model to estimate the impact of air pollution on housing prices. Using the natural logarithm of housing prices as the dependent variable, PM2.5 as the independent variable, adding control variables and plant-year-month fixed effects. The results indicate that a one unit (ug/m³) increase in PM2.5 is associated with a statistically significant decrease of 3.33% on housing prices. Based on an analysis of data on housing, PM2.5, and Distinct Angle IV, we have shown again about the negative relationship between air pollution and housing prices.

The time-lag effects are a universal phenomenon within the sphere of economic activity. Considering the enduring nature of housing buying decisions, as opposed to being impulsive in nature, it is widely accepted that housing prices is not only influenced by present-day air quality but also account for the preceding period's air pollution conditions in the housing location. Column (2) of Table 7 reports the results of a robustness test, in which the PM2.5 and instrumental variables were replaced with their respective values from the previous period. The analysis shows that the housing prices are highly responsive to PM2.5, with an estimated coefficient of approximately -1.90%. Furthermore, our results reveal that the direction, statistical significance, and magnitude of the estimated coefficients closely correspond to the marginal effects derived from the baseline regression analysis, thereby emphasizing the validity of our findings.

In Column (3), the dependent variable has been modified from housing transaction prices to housing listing prices. Correspondingly, PM2.5 and instrumental variables have been substituted with the relevant values of the listing period. As shown in Column (3), Distinct Angle IV method provides evidence that air pollution constitutes a significant factor in the determination of housing listing price. These findings suggest that PM2.5 has a dual effect, not only reflect in the preferences and willingness to pay of prospective households, but also in the price expectations of sellers.

Column (4) performs a robustness check by screening the housing samples with the restriction that there is only one thermal power plant within 6km of the house. Using Equation (6) to perform regression, the results in Column (4) show that 1% increase in PM2.5 is associated with a decrease in housing prices by 2.34%. The elasticity of housing prices with respect to air pollution obtained in the restricted regression is like that estimated in the baseline regression (-2.07%), implying that our results are robust.

Columns (5)-(6) of Table 7 present the results of the second stage of Distinct Angle IV method, with a slight adjustment for the distance constraint between the house and its nearest thermal power plant. The estimated elasticities of housing prices with respect to PM2.5 are observed to be -2.38% and -2.39%, respectively, which does not significantly diverge in magnitude from the original regression results (-2.07%). Based on these estimated elasticities show a high degree of similarity, thereby attesting to the robustness of our findings against variations in the maximum distance radius defining the housing sub-market.

In addition, the five robustness test methods in Table 7 all reject the hypothesis that Distinct Angle IV is a weak instrumental variable. Thus, it can be judged that we have effectively solved the endogeneity problem.

Table 7: IV results from the second step estimation: robustness checks

Ln(Price)	Ln(Price)	Ln(listing-price)	Ln(Price)	Ln(Price)	Ln(Price)
(1)	(2)	(3)	(4)	(5)	(6)
A semi-log model	Previous period	Replace the explained variable	select subsample	Radius of 5.5km	Radius of 6.5km

PM2.5	-0.033***					
	(0.009)					
Ln(PM2.5)		-1.897***	-4.523***	-2.334***	-2.381***	-2.387***
		(0.582)	(1.424)	(0.751)	(0.661)	(0.626)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Plant*Year*Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	583,021	582,849	526,388	392,860	527,798	640,652
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000
Kleibergen-Paap Wald						
rk F statistic	41.32	34.58	16.26	26.46	32.81	33.41

Notes:

(1) Table 7 presents the results of five different robustness tests. Column (1) uses a semi-log model to examine the impact of unit change in PM2.5 on housing prices. Column (2) regresses current housing prices using time-varying variables (e.g., PM2.5) in the previous period and other time-invariant variables. Column (3) uses the method of replacing the dependent variable with housing listing-prices. Column (4) eliminates the potential interference caused the presence of multiple thermal power plants around houses. Column (5)-(6) try to use different thresholds to construct instrumental variables.

(2) Table 7 uses Distinct Angle IV method to address the endogeneity problem of air pollution. All regressions add control variables and plant-year-month fixed effects. Standard errors are clustered at the plant-year-month level. Standard errors in parentheses, *p<0.1, **p<0.05, ***p<0.01.

VI.E The influence of public service facilities on the negative impact of air pollution

After considering estimations bias from migration costs, endogenous problems and the government's price control policy, the estimated implicit price of air quality is credible. Based on above estimation method, this section analyzes the impact of government behavior on negative externalities of air pollution from the perspective of government investment. Empirical evidence suggests that the implementation of a policy targeted at reducing air pollution can lead to a decrease in the negative externality to society, while efforts to increase awareness of the negative impact of pollution can enhance household's' marginal willingness to pay. However, empirical evidence is lacking in concerning the impact of government investment on the housing market's response to air pollution. This study analysis the influence of public service facilities on the negative impact of air pollution by means of sub-sample regression, which makes up for the above shortcomings.

First, we estimate the influence of transportation facilities on the negative impact of air

pollution, by comparing the elasticity of housing prices with respect to PM2.5 between houses with subway stations within 0.5km and those without. Columns (1)-(2) in Table 8 show that the presence of a subway station will lead to a change in the elasticity from a significant -2.67% to a not statistically significant -0.94%. There is a reduction in marginal willing to pay for clean air for households near subway stations as compared with households far away from subway stations, which may be caused by differences in transportation convenience.

Second, Columns (3)-(4) in Table 8 estimate the influence of environmental facilities on the negative impact of air pollution. We set up a 0-1 dummy variable based on whether there is a park within 0.5km of the house, and thus dividing the housing samples into two categories. The results show that the elasticity of housing prices with respect to air pollution will decrease from 3.24% to 0.92% if there is existing a park. This may be because, on the one hand, parks can improve the air environment of the city. When there is a park, people's expectations for the local air environment will be better. On the other hand, environmental facilities like parks are powerful measures to improve living comfort. Parks cannot only meet people's health needs, but also satisfy their needs for physical and mental relaxation. Therefore, households living near the park will have a lower marginal willing to pay for clean air, because when they move to places far away from the park due to the improvement of the air environment, they will lose the living comfort brought by the park.

Table 8: The impact of transportation and environmental public service facilities

	(1)	(2)	(3)	(4)
	Subway		Park	
	With	Without	With	Without
Ln(PM2.5)	-0.940	-2.662***	-0.920*	-3.243***
	(0.610)	(0.721)	(0.545)	(0.858)
Controls	Yes	Yes	Yes	Yes
Plant*Year*Month FE	Yes	Yes	Yes	Yes
<i>N</i>	167,828	414,673	268,898	313,014
Prob > F	0.000	0.000	0.000	0.000
Kleibergen-Paap Wald rk F statistic	26.17	31.94	33.27	28.03

Notes: Table 8 uses Distinct Angle IV method to address the endogeneity problem of air pollution. All regressions take the natural logarithm of housing prices as the dependent variable, and the natural logarithm of PM2.5 as the independent variable, adding control variables and plant-year-month fixed effects. Standard errors are clustered at the plant-year-month level. Standard errors in parentheses, *p<0.1, **p<0.05, ***p<0.01.

Similarly, Table 9 divides the samples into sub-samples according to the presence of a primary school or general hospital within a 0.5-kilometer radius from the house. The empirical results show that education facilities and medical facilities provided by the government can effectively mitigate the negative impact of air pollution by decreasing the households' sensitivity to air pollution.

Our findings support to the proposition that the provision of public service facilities can obviously impact the implicit price of air quality in adjacent areas. This means that governments have a crucial role to play in addressing the negative impact of air pollution on social wellbeing. Public service facilities can effectively serve as a form of compensation for households living in heavily polluted areas. The government possess the capability to mitigate the negative impact of air pollution by expanding public service facilities, not just limited to the traditional total pollution control policy.

Table 9: The impact of education and medical public service facilities

	(1)		(2)		(3)		(4)	
	School				Hospital			
	With	Without	With	Without	With	Without	With	Without
Ln(PM2.5)	-1.274**	-3.149***	-0.603	-4.704***				
	(0.511)	(0.945)	(0.388)	(1.356)				
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant*Year*Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	319,637	262,261	284,903	297,008				
Prob > F	0.000	0.000	0.00	0.000				
Kleibergen-Paap Wald rk F statistic	28.17	30.45	46.67	19.93				

Notes: Table 9 uses Distinct Angle IV method to address the endogeneity problem of air pollution. All regressions take the natural logarithm of housing prices as the dependent variable, and the natural logarithm of PM2.5 as the independent variable, adding control variables and plant-year-month fixed effects. Standard errors are clustered at the plant-year-month level. Standard errors in parentheses, *p<0.1, **p<0.05, ***p<0.01.

VII. Conclusion

Air pollution is widely believed to cause healthy risks, social welfare and economy loose. This study estimates the impact of air pollution on housing prices using the hedonic price model. A unique strength of this study is the construction of a dataset that cleanly identifies causal effects. The unique data set constructed through the integration of second-hand housing data, air pollution satellite-based data, thermal power plants data, meteorological data, POI data, and VIIRS nighttime light data, all linked through geographical location information.

We provide credible evidence on households' marginal willingness to pay for clean air in China, a developing countries, after solving three problems of migrations costs, endogenous problems of air pollution and the government's price control policy. Our instruments weaken the bias of the hedonic estimates, estimated results show that the hedonic estimation obtained without paying attention to these three problems is downward biased. The estimated elasticity of housing prices with respect to PM2.5 combined with Distinct Angle IV method shows that, on average, every 1% increase in PM2.5 leads to a 2.07% reduction in housing prices. In contrast, after controlling for the plant-year-month fixed effects and housing characteristics, OLS method yields a positive impact of PM2.5 on housing prices, which is odds at intuition. These results have important implications for policy makers, suggesting that the economic benefits of regulations that reduced air pollution are substantially larger than those found in previous studies which have ignored migration costs. The welfare effects of environmental quality improvements are substantial in China.

More broadly, we think this study has many implications for future research. First, the findings suggest that the housing market can be used to determine the value of environmental facilities, as well as to determine the value of non-market goods. Second, endogenous problems are often faced by air pollution research. We show that wind direction is a good instrumental variable for air pollution, thereby addressing the mutual endogenous of air pollution and housing prices. Furthermore, by segmenting multiple housing markets, it is possible to mitigate migration costs and enhance identification of cleaner estimates, thereby achieving greater accuracy in the quantification of housing market effects.

References

Auffhammer, M., Wang, M., Xie, L., & Xu, J. (2021). Renewable electricity development in China: Policies, performance, and challenges. *Review of Environmental Economics and Policy*, 15(2), 323-339.

Banzhaf H S. Difference-in-differences hedonics[J]. *Journal of Political Economy*, 2021, 129(8): 2385-2414.

Bayer P, Keohane N, Timmins C. Migration and hedonic valuation: The case of air quality[J]. *Journal of Environmental Economics and Management*, 2009, 58(1): 1-14.

Chang T Y, Graff Zivin J, Gross T, et al. The effect of pollution on worker productivity: evidence from call center workers in China[J]. *American Economic Journal: Applied Economics*, 2019, 11(1): 151-72.

Chay K Y, Greenstone M. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession[J]. *The quarterly journal of economics*, 2003, 118(3): 1121-1167.

Chay K Y, Greenstone M. Does air quality matter? Evidence from the housing market[J]. *Journal of political Economy*, 2005, 113(2): 376-424.

Chen S, Jin H. Pricing for the clean air: Evidence from Chinese housing market[J]. *Journal of Cleaner Production*, 2019, 206: 297-306.

Chen S, Oliva P, Zhang P. Air pollution and mental health: evidence from China[R]. National Bureau of Economic Research, 2018.

Chen S, Oliva P, Zhang P. The effect of air pollution on migration: evidence from China[J]. *Journal of Development Economics*, 2022, 156: 102833.

Chen Y, Ebenstein A, Greenstone M, et al. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy[J]. *Proceedings of the National Academy of Sciences*, 2013, 110(32): 12936-12941.

Currie J, Davis L, Greenstone M, et al. Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings[J]. *American Economic Review*, 2015, 105(2): 678-709.

Deryugina T, Heutel G, Miller N H, et al. The mortality and medical costs of air pollution: Evidence from changes in wind direction[J]. *American Economic Review*, 2019, 109(12): 4178-4219.

Deschenes O, Greenstone M, Shapiro J S. Defensive investments and the demand for air quality: Evidence from the NOx budget program[J]. *American Economic Review*, 2017, 107(10): 2958-2989.

Dominici F, Greenstone M, Sunstein C R. Particulate matter matters[J]. *Science*, 2014, 344(6181): 257–259.

Ebenstein A, Fan M, Greenstone M, et al. New evidence on the impact of sustained exposure to air pollution on life expectancy from China’s Huai River Policy[J]. *Proceedings of the National Academy of Sciences*, 2017, 114(39): 10384-10389.

Ebenstein A, Lavy V, Roth S. The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution[J]. *American Economic Journal: Applied Economics*, 2016, 8(4): 36-65.

Elvidge C D, Baugh K, Zhizhin M, et al. VIIRS night-time lights[J]. *International journal of remote sensing*, 2017, 38(21): 5860-5879.

Freeman R, Liang W, Song R, et al. Willingness to pay for clean air in China[J]. *Journal of Environmental Economics and Management*, 2019, 94: 188-216.

Fu S, Viard V B, Zhang P. Air pollution and manufacturing firm productivity: Nationwide estimates for China[J]. *The Economic Journal*, 2021, 131(640): 3241-3273.

Grainger C A. The distributional effects of pollution regulations: Do renters fully pay for cleaner air?[J]. *Journal of Public Economics*, 2012, 96(9-10): 840-852.

Greenstone M, He G, Li S, et al. China’s war on pollution: Evidence from the first 5 years[J]. *Review of Environmental Economics and Policy*, 2021, 15(2): 281-299.

Greenstone M, Jack B K. Envirodevonomics: A research agenda for an emerging field[J]. *Journal of Economic Literature*, 2015, 53(1): 5-42.

Gupta P, Sundar A C, Jun W, et al. Satellite Remote Sensing of Particulate Matter and Air Quality Assessment over Global Cities[J]. *Atmospheric Environment*, 2006, 40(30): 5880–5892.

Han X, Zhao B. Luxury Blue Sky-the Estimation of Air Quality Premium in Housing Prices and Its Heterogeneity[J]. *China Economics Quarterly*, 2021,21(03):755-774.

Hanna R, Oliva P. The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City[J]. *Journal of Public Economics*, 2015, 122: 68-79.

He J, Liu H, Salvo A. Severe air pollution and labor productivity: Evidence from industrial towns in China[J]. *American Economic Journal: Applied Economics*, 2019, 11(1): 173-201.

Heyes A, Zhu M. Air pollution as a cause of sleeplessness: Social media evidence from a panel of Chinese cities[J]. *Journal of Environmental Economics and Management*, 2019, 98: 102247.

Ito K, Zhang S. Willingness to pay for clean air: Evidence from air purifier markets in China[J]. *Journal of Political Economy*, 2020, 128(5): 1627-1672.

Karplus V J, Zhang J, Zhao J. Navigating and evaluating the labyrinth of environmental regulation in China[J]. *Review of Environmental Economics and Policy*, 2021, 15(2): 300-322.

Kumar N, Chu A D, Foster A D, et al. Satellite Remote Sensing for Developing Time and Space Resolved Estimates of Ambient Particulate in Cleveland, OH[J]. *Aerosol Science and Technology*, 2011, 45(9):1090–1108.

Qin Y, Zhu H. Run away? Air pollution and emigration interests in China[J]. *Journal of Population Economics*, 2018, 31(1): 235-266.

Rosen S. Hedonic prices and implicit markets: product differentiation in pure competition[J]. *Journal of political economy*, 1974, 82(1): 34-55.

Shi K, Yu B, Huang Y, et al. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data[J]. *Remote Sensing*, 2014, 6(2): 1705-1724.

Smith V K, Huang J C. Can markets value air quality? A meta-analysis of hedonic property value models[J]. *Journal of political economy*, 1995, 103(1): 209-227.

Tiebout C M. A pure theory of local expenditures[J]. *Journal of political economy*, 1956, 64(5): 416-424.

Van Donkelaar A, Hammer M S, Bindle L, et al. Monthly global estimates of fine particulate matter and their uncertainty[J]. *Environmental Science & Technology*, 2021, 55(22): 15287-15300.

Yusuf A A, Resosudarmo B P. Does clean air matter in developing countries' megacities? A hedonic price analysis of the Jakarta housing market, Indonesia[J]. *Ecological Economics*, 2009, 68(5): 1398-1407.

Zhang X, Chen X, Zhang X. The impact of exposure to air pollution on cognitive performance[J]. *Proceedings of the National Academy of Sciences*, 2018, 115(37): 9193-9197.

Zhang J, Mu Q. Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks[J]. *Journal of Environmental Economics and Management*, 2018, 92: 517-536.

Zhang S, Zhou Y, Xu P. Air quality affects house prices—Analysis based on RD of the Huai River policy[J]. *Sustainable Cities and Society*, 2022, 85: 104017.

Zheng Q, Deng J, Jiang R, et al. Monitoring and assessing “ghost cities” in Northeast China from the view of nighttime light remote sensing data[J]. *Habitat International*, 2017, 70: 34-42.

Zheng S, Kahn M E. A new era of pollution progress in urban China?[J]. *Journal of Economic Perspectives*, 2017, 31(1): 71-92.

Zivin J G, Neidell M. Days of haze: Environmental information disclosure and intertemporal avoidance behavior[J]. *Journal of Environmental Economics and Management*, 2009, 58(2): 119-128.

Zivin J G, Neidell M. The impact of pollution on worker productivity[J]. *American Economic Review*, 2012, 102(7): 3652-3673.

Appendix

A: The specific use of instrumental variables

Distinct Angle IV is a continuum, whose value ranges from 0 to 180.

45°interval Angle IV divides Monthly Angle into intervals of 45 degrees. In accordance with this IV method, four dummy variables are used as demonstrated in Equation (A1):

$$\left\{ \begin{array}{l} Angle[0,45) = \begin{cases} 1, & \text{if } 0^\circ \leq Angle < 45^\circ \\ 0, & \text{others} \end{cases} \\ Angle[45,90) = \begin{cases} 1, & \text{if } 45^\circ \leq Angle < 90^\circ \\ 0, & \text{others} \end{cases} \\ Angle[90,135) = \begin{cases} 1, & \text{if } 90^\circ \leq Angle < 135^\circ \\ 0, & \text{others} \end{cases} \\ Angle[135,180) = \begin{cases} 1, & \text{if } 135^\circ \leq Angle < 180^\circ \\ 0, & \text{others} \end{cases} \end{array} \right. \quad (A1)$$

60°interval Angle IV divides Monthly Angle into intervals of 60 degrees. In accordance with this IV method, three dummy variables are used as demonstrated in Equation (A2):

$$\left\{ \begin{array}{l} Angle[0,60) = \begin{cases} 1, & \text{if } 0^\circ \leq Angle < 60^\circ \\ 0, & \text{others} \end{cases} \\ Angle[60,120) = \begin{cases} 1, & \text{if } 60^\circ \leq Angle < 120^\circ \\ 0, & \text{others} \end{cases} \\ Angle[120,180) = \begin{cases} 1, & \text{if } 120^\circ \leq Angle < 180^\circ \\ 0, & \text{others} \end{cases} \end{array} \right. \quad (A2)$$

45°Bins Angle IV method considers the difference in frequency of the prevailing wind direction. For example, in extreme cases, a thermal power plant blows north wind for 18 days and south wind for 12 days in one month; and in another month blows north wind for 28 days and departs from the south for 2 days. Although the monthly prevailing wind direction obtained for the two months are the same, the air pollution exposures caused by the two wind frequency combinations are significantly different. In accordance with this IV method, four dummy variables are used as demonstrated in Equation (A3):

$$45^\circ\text{Bins} \left\{ \begin{array}{l} Angle[0,45) = \text{Total days that Angle falls in } 0\sim 45 \text{ interval in a month} \\ Angle[45,90) = \text{Total days that Angle falls in } 45\sim 90 \text{ interval in a month} \\ Angle[90,135) = \text{Total days that Angle falls in } 90\sim 135 \text{ interval in a month} \\ Angle[135,180) = \text{Total days that Angle falls in } 135\sim 180 \text{ interval in a month} \end{array} \right. \quad (A3)$$

In accordance with 60°Bins Angle IV method, three dummy variables are used as demonstrated in Equation (A4):

$$60^\circ\text{Bins} \begin{cases} \text{Angle}[0,60) = \text{Total days that Angle falls in } 0\sim 60 \text{ interval in a month} \\ \text{Angle}[60,120) = \text{Total days that Angle falls in } 60\sim 120 \text{ interval in a month} \\ \text{Angle}[120,180) = \text{Total days that Angle falls in } 120\sim 180 \text{ interval in a month} \end{cases} \quad (\text{A4})$$

B: Application of the unit vector method

The unit vector method is expressed as follows:

$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^n \sin A_i \quad (\text{B1})$$

$$\bar{\nu} = \frac{1}{n} \sum_{i=1}^n \cos A_i \quad (\text{B2})$$

$$A = \arctan \left(\frac{\bar{\mu}}{\bar{\nu}} \right) \quad (\text{B3})$$

The wind direction can be quantified using three variables: $\bar{\mu}$, $\bar{\nu}$ and A . $\bar{\mu}$ represents the average east-west component. $\bar{\nu}$ represents the mean north-south component. A represents the mean wind direction calculated using the unit vector method. The calculation of the average wind direction using the unit vector method involves projecting the wind direction scale onto the X-Y coordinate axis at each statistical moment. The cumulative projection average is then calculated, and the true average angle is restored using the arctangent function. However, there is a problem that needs to be considered during the data processing stage. The arctangent function has a value orientation of -90° to 90° , but the actual wind direction angle has a range of 0° to 360° . As a result, half of the angles are not representable. Therefore, we divide the angles according to the quadrants in mathematics, the first step is judging the quadrant where the real angle is located by the positive and negative values of the east-west azimuth average component $\bar{\mu}$ and the north-south azimuth average component $\bar{\nu}$, and then perform angle correction according to the quadrant attributes, and finally obtain the formal average wind direction. In particular:

$$\text{correction value} = \begin{cases} +0^\circ, \text{ if } \bar{\mu} > 0 \text{ and } \bar{\nu} > 0 \\ +180^\circ, \text{ if } \bar{\mu} > 0 \text{ and } \bar{\nu} < 0 \\ +180^\circ, \text{ if } \bar{\mu} < 0 \text{ and } \bar{\nu} < 0 \\ +360^\circ, \text{ if } \bar{\mu} < 0 \text{ and } \bar{\nu} > 0 \end{cases} \quad (\text{C4})$$

