PHBS WORKING PAPER SERIES

The Role of Information in the Rosen-Roback Framework

Xuwen Gao*
Peking University

Ran Song
Yale-NUS College

Christopher Timmins
Duke University and NBER

May 2021

Working Paper 20210502

Abstract

In this paper, we study the role of information in non-market valuation. We develop a variant of the Rosen-Roback model of inter-urban sorting that incorporates public access to information about air quality, and demonstrate that information constraints create a wedge between the revealed and true hedonic prices. Moreover, the direction and the magnitude of that wedge depends on the individual's perception bias. We empirically test our theoretical predictions by leveraging a natural experiment – the unexpected disclosure of PM2.5 data in China. We find that perception bias before the data disclosure leads to a downward estimation bias in hedonic valuation. The hedonic price of avoiding PM2.5 exposure increases substantially from 169 to 337 Chinese Yuan in response to the information shock. Our work sheds light on the application of Rosen-Roback theory in countries where the public access to information is restricted.

Keywords: Information, Rosen Roback Theory, wage-hedonic model, compensating differential
JEL Classification: Q51, Q53, R23

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China

The Role of Information in the Rosen-Roback Framework*

Results are preliminary and subject to change

Xuwen Gao

Peking University

Ran Song

Yale-NUS College

Christopher Timmins

Duke University and NBER

Abstract

In this paper, we study the role of information in non-market valuation. We develop a variant of the Rosen-Roback model of inter-urban sorting that incorporates public access to information about air quality, and demonstrate that information constraints create a wedge between the revealed and true hedonic prices. Moreover, the direction and the magnitude of that wedge depends on the individual's perception bias. We empirically test our theoretical predictions by leveraging a natural experiment – the unexpected disclosure of PM2.5 data in China. We find that perception bias before the data disclosure leads to a downward estimation bias in hedonic valuation. The hedonic price of avoiding PM2.5 exposure increases substantially from 169 to 337 Chinese Yuan in response to the information shock. Our work sheds light on the application of Rosen-Roback theory in countries where the public access to information is restricted.

JEL Classification: Q51, Q53, D83

Keywords: Information, Rosen-Roback Theory, Wage-hedonic model, Compensating differential

---

*Xuwen Gao, Peking University HSBC Business School, Shenzhen 510632, China, Email: gaoxw@phbs.pku.edu.cn. Ran Song, Division of Social Science, Yale-NUS College, 138527, Singapore, Email: ran.song@yale-nus.edu.sg. Christopher Timmins, Department of Economics, Duke University and NBER, Durham 27708, United States, Email: christopher.timmins@duke.edu. (*Corresponding Author*). Ran Song acknowledges the research support from Yale-NUS College and the Tan Chin Tuan Chinese Culture Civilization Program in Singapore.

1. Introduction

In the past forty years, the Rosen-Roback hedonic model (Rosen, 1979; Roback, 1982) has become an important tool for valuing non-market amenities. An underlying assumption of the Rosen-Roback hedonic framework is that households have complete information about the level of amenities across space. On the basis of this assumption, households sort across locations reflecting the tradeoff between income, housing costs and amenities. Therefore, the hedonic price measured by geographical differences in income and housing prices reflects the implicit value of amenities such as environmental quality.

However, this complete information assumption often fails to hold. In many highly polluted developing countries, real-time environmental information is either blocked by policy makers or is never even gathered. Even in some developed countries, pollution monitoring networks are sparse, failing to cover a large fraction of population. Information constraints could prevent amenity-induced spatial sorting, and create a wedge between the estimated hedonic price and the real implicit value of pollution abatement. As a consequence, researchers may recover biased estimates of individuals' Marginal Willingnesses to Pay (MWTP) for improvement to environmental quality when relying on observed residential decisions.

In this paper, we relax the assumption of complete information in the Rosen-Roback model. We theoretically analyze the role of information in non-market valuation in a simple equilibrium framework, and then leverage the unexpected disclosure of PM2.5 data in China to evaluate the causal association between information availability and the hedonic price of air quality.

We first establish a simple variant of Rosen-Roback model incorporating the public access to information. At its heart is the idea that people maximize utility via their residential sorting decisions by trading off income, housing costs, and their *perceived level* of a location specific amenity. In equilibrium under complete information, income and house price gradients will capture the implicit value of amenities like air quality. However, in equilibrium under incomplete information, there will exist a wedge between regional differences in income and housing prices and the true value of the amenity. Information constraints bias hedonic price estimates, and the magnitude and the direction of that bias will depend upon the magnitude and the direction of decision-makers' perception biases. Our adaptation of the Rosen-Roback model quantifies the implications of information failure on hedonic values.

China provides an ideal laboratory to test the predictions of our modified Rosen-Roback model. Despite being exposed to hazardous levels of pollution, Chinese citizens used to have limited or no access to information about local air quality. In 2013, China launched a nation-wide, real-time air quality monitoring and disclosure program, incorporating PM2.5 concentration into the air quality measure for the first time. The information shock triggered the spatial sorting of people in China (Khanna, Liang, Song and Mobarak, 2021) and allows us to quantify changes in the compensating differential for pollution driven by the provision of pollution information.

We derive a wage hedonic equation from our equilibrium framework with information constraints, in which the outcome variable is income adjusted by local housing prices.[1] Thus, our hedonic price measure reflects spatial differences in both income and housing costs. Our primary variable of interest is the interaction of an information disclosure indicator for the level of pollution, which represents the wedge between the revealed hedonic price (under incomplete information) and implicit value of clean air in the Rosen-Roback Framework.

We leverage the staggered publication of the real time PM2.5 data across three waves of Chinese cities. The central government determined the sequence of staggered rollout according to the location and the tier of cities and several pre-determined designations. Since these conditions were established prior to the information program, controlling for city-fixed effects allows us to account for unobservables that might affect the sequence of the data rollout.

We isolate exogenous fluctuations in pollution, leveraging variation in wind direction combined with the historical placement of distant thermal power plants (as in Freeman et al., 2019), and a regression discontinuity around the Huai river (as in Chen et al., 2013). We estimate the interaction between the information disclosure indicator and the level of pollution, controlling simultaneously for the two independent components of the interaction term. Adding the information disclosure indicator could account for remaining unobservables that may be correlated with the program, and controlling for the level of pollution could disentangle remaining unknown confounding factors that our instruments fail to address.

In order to measure air quality before and after China's disclosure of official PM2.5 data, we employ satellite-based PM2.5 data measured using the Global Annual PM2.5 Grids derived by Van

[1] In our paper, we consider the geographical variation in both nominal income and housing prices driven by amenity differences. Therefore, throughout our paper, we define the compensating differential as the additional amount of real income (nominal income deflated by the housing price) that an individual must be offered in order to motivate them to accept a reduction in the level of amenity.

Donkelaar et al. (2016). The data have a high grid cell resolution of 0.01 degree, and can provide a comprehensive and reliable measurement of air quality for all Chinese cities. The correlation between satellite-based PM2.5 data and monitor-based PM2.5 data in China is up to 0.8 (Freeman, Liang, Song, Timmins, 2019).

We find salient differences in the compensating differential for PM2.5 pollution under heterogenous scenarios of information availability. The disclosure of pollution data significantly increases the hedonic price of avoiding PM2.5 exposure in China. Driven by the unexpected dissemination of information, a median individual's MWTP for a one-unit reduction in PM2.5 concentration raises from 169 Chinese Yuan under *incomplete* information to 337 Chinese Yuan under *complete* information. This increase represents approximately 0.84% of the median individual's income. Our research confirms a news report saying that firms in China pay substantial 'pollution premiums' to attract workers after pollution information was widely disseminated (New York Times, 2015).

Our results indicate that information constraints lead to a wedge between the revealed hedonic price and the true implicit value of clean air; the magnitude of the wedge accounts for as large as about 17.5% of the implicit value. These findings verify the predictions of our theoretical framework and demonstrate the importance of incorporating access to information in the Rosen-Roback hedonic model. As Chinese people tend to underestimate the level of air pollution before the data disclosure (Barwick et al., 2019), the downward perception bias leads to a downward estimation bias in hedonic valuation. Overlooking the role of information would severely understate the economic value of amenities and undermine benefit-cost analysis for important public policies.

Our primary contribution lies in the incorporation of public access to information into the Rosen-Roback hedonic theory and the analysis of the implications of information constraints on the applicability of the hedonic methodology. Under incomplete information, people would maximize their *perceived* utility based on their *perceived* level of amenities, which they would trade-off against earnings and housing prices across space. Perception bias would distort the relationship between the real implicit value and the revealed hedonic price of amenities, biasing MWTP estimates. We are among the first to analyze the role of information in the Rosen-Roback framework, and to both theoretically and empirically demonstrate how information failure affects the foundation of revealed-preference framework.[2]

We also speak to the literature in applied economics that employs the hedonic approach to measure compensating differentials. Environmental and urban economists calculate the economic value of (dis)amenities based on revealed compensating differentials (Blomquist et al. 1988), and labor economists use compensating differentials to measure the economic value of job attributes (Hersch, 1998) and the value of statistical life (Viscusi, 1993; Aldy and Taylor, 2019). All these papers implicitly assume that people have complete information on these variables, and we are unaware of any paper evaluating the association between information integrity and revealed compensating differentials. Against this backdrop, we analyze how the access to information determines the validity of the compensating differential estimates.

Our work sheds light on the measure of city quality of life (QOL). A large body of literature measures the QOL in the U.S, based on differentials in income and housing costs across cities (Rosen, 1979; Roback, 1982; Kahn, 1995; Albouy, 2008). However, econometricians cannot simply borrow the same methodology to estimate the QOL in developing counties, where official data on important city attributes – e.g., environmental pollution or the crime rate – are not available or may be manipulated. Incomplete information would differentiate the ranking of estimated city QOL from true ranking of city livability.

The empirical part of our work builds on contemporaneous works on willingness to pay for clean air (Freeman, Liang, Song, Timmins, 2019; Ito and Zhang, 2020) and the economic effects of information disclosure (Barwick et al., 2019; Wang and Zhang, 2021) in China.[3] Barwick et al. (2019) conduct pioneering research on the effects of pollution information on a wide range of avoidance behaviors and health outcomes. They use AOD concentration as the pollution measure and also explore the exogeneous shock of PM2.5 data disclosure in China. In contrast, our empirical application uses satellite PM2.5 data that are highly consistent with the official PM2.5 measure and are available both before and after the release of official data, and analyze the how

[2] An alternative widely used revealed-preference methodology is the general equilibrium sorting model. Although we focus on how information failure biases the MWTP estimates from the hedonic model, our theoretical analysis indicates that incomplete information also undermines the estimation of the MWTP via equilibrium soring model.

[3] Freeman, Liang, Song and Timmins (2019) estimate the willingness to pay for clean air based on the spatial variation in income and housing prices. Ito and Zhang (2020) examine air purifier transactions to recover household preference for the reduction in indoor PM10 concentration. Wang and Zhang (2021) study the effects of information disclosure on face mask consumption in China.

restrictions on the access to PM2.5 information biases the hedonic price of avoiding PM2.5 exposure estimated by regional differences in income and housing prices.

The remainder of this paper proceeds as follows. Section 2 lays out our adaption of the Rosen-Roback model to consider information constraints. Section 3 discusses the natural experiment of data disclosure in China, and Section 4 describes the data. Section 5 presents estimation techniques used to test our theoretical predictions and Section 6 reports estimation results. Section 7 concludes.

2. A Rosen-Roback Framework with Incomplete Information

We relax the underlying assumption of complete information and build a variant of Rosen-Roback model incorporating the access to amenity information. Under complete information, people maximize their utility via geographical sorting, arbitraging location specific amenities, income, and housing costs. Thus, in equilibrium, spatial differences in income and housing prices reflect the implicit value of the amenity. However, under incomplete information, the public tends to misinterpret the level of the amenity, distorting the arbitrage process. As a result, incomplete information leads to a wedge between income and housing price differentials and the true value of the amenity, biasing hedonic estimates.

2.1 The Hedonic Price under Complete Information

In this section, we develop a simple version of the Rosen-Roback model with complete information to use as a starting point. Each individual's residential location choice set is characterized by a location-specific amenity (say, air quality). When complete information on the amenity is available, individual i chooses residential city j to maximize its indirect utility:

$$V_{i,j} = I_{i,j}\rho_j^{-\theta} X_j^{\gamma} \tag{1}$$

where $I_{i,j}$ is the nominal income that individual i could earn in city j, and ρ_j and X_j represent the unit price of housing services and the amount of the amenity in city j, respectively. θ is the fraction of housing expenditure in income, and γ measures preferences for the components of X_j.

Under complete information, individual i makes a tradeoff between amenities, income and housing costs via spatial sorting, choosing the optimal level of amenity to maximize their utility. The core of the hedonic framework requires that, in equilibrium, identical individuals must be indifferent among locations; if not, movements would occur to arbitrage away utility differences.

Holding indirect utility as fixed at some common level $\bar{V}$, we can take the total derivative with respect to the level of amenity:

$$\frac{\partial V_{i,j}}{\partial X_j} = V_{I_j} * \frac{\partial I_{i,j}}{\partial X_j} + V_{\rho_j} * \frac{\partial \rho_j}{\partial X_j} + V_{X_j} = 0 \tag{2}$$

Using Roy's identity, the optimum consumption of housing services for individual i in city j is:

$$H_{i,j}^{*} = -\frac{V_{\rho_j}}{V_{I_{i,j}}} = \frac{\theta I_{i,j}}{\rho_j} \tag{3}$$

Substituting equation (3) into (2), the marginal willingness to pay for X with complete information ($MWTP_X^C$) for changes in the amenity X_j can be written as the marginal rate of substitution between income and the amenity:

$$MWTP_X^C = \frac{V_{X_j}}{V_{I_{i,j}}} = \left(\frac{\partial \log I_{i,j}}{\partial X_j} - \theta \frac{\partial \log \rho_j}{\partial X_j}\right) \times I_{i,j} = \gamma \left(\frac{I_{i,j}}{X_j}\right) \tag{4}$$

Equation (4) is widely used for non-market valuation in literature. In equilibrium under complete information, the MWTP for X_j is determined by the preference parameter γ, and can be fully captured by regional differences in income and housing prices. Therefore, econometricians can rely on income and housing price gradient to calculate the implicit value of the amenity.

2.2 The Hedonic Price under Incomplete Information

When people do not have full access to information about amenities, the level of the location-specific amenity perceived by individuals is $\widetilde{X_J}$:

$$\widetilde{X_J} = X_j^{\lambda} \tag{3}$$

where λ measures the magnitude and the direction of perceptual bias. If people tend to understate the actual level of the amenity, $\lambda < 1$; if they overstate the level of the amenity, $\lambda > 1$. Individual i will then sort across locations in order to maximize their *perceived* indirect utility:

$$\widetilde{V_{i,J}} = I_{i,j} \rho_j^{\theta} X_j^{-\lambda\gamma} \tag{4}$$

Returning to sorting equilibrium and taking the total derivative of perceived indirect utility with respect to the level of amenity yields:

$$\frac{\partial \tilde{V}_{i,j}}{\partial X_j} = \tilde{V}_{I_j} * \frac{\partial I_{i,j}}{\partial X_j} + \tilde{V}_{\rho_j} * \frac{\partial \rho_j}{\partial X_j} + \tilde{V}_{X_j} = 0 \tag{5}$$

The optimal consumption of housing services conditional upon choosing a particular city remains the same as equation (3). Thus, the marginal willingness to pay for X_j under incomplete information ($MWTP_X^I$) is:

$$MWTP_X^I = \frac{\tilde{V}_{X_j}}{\tilde{V}_{I_j}} = \left(\frac{\partial \, log \, I_{i,j}}{\partial X_j} - \theta \frac{\partial \, log \, \rho_j}{\partial X_j}\right) \times I_{i,j} = \lambda\gamma \left(\frac{I_{i,j}}{X_j}\right) \tag{7}$$

MWTP differs between scenarios of complete and incomplete information:

$$MWTP_X^I = \lambda\gamma \left(\frac{I_{i,j}}{X_j}\right) \neq MWTP_X^C = \gamma \left(\frac{I_{i,j}}{X_j}\right) \tag{8}$$

and $MWTP_X^I$ will not represent the actual preference for X_j due to the perception bias captured by λ. The researcher cannot therefore rely on spatial differences in income and housing prices to quantify the implicit value of X_j. Information constraints could explain the unreasonable low MWTP for environmental quality improvement in many highly polluted developing countries (an issue raised by Greenstone and Jack (2015))[4].

3. Natural Experiment in China

China provides a unique setting in which to study the impacts of information constraints on hedonic valuation. Despite the hazardous level of exposure to pollution, Chinese citizens used to have limited or no access to real time information about local air quality. In 2000, China started to publish air quality data, including an Air Pollution Index (API) along with specific information on PM10,[5] but only did so for 42 cities. The number of cities in which this information was available increased gradually to 120 in 2012. However, official API and PM10 data were vulnerable to the manipulation of local government prior to 2012 due to weak monitoring and enforcement of the central government (Chen et al., 2012; Ghanem and Zhang, 2014).

While air pollution information prior to 2012 focused on SO2, NO2 and PM10, fine particulate pollution has been the most important source of air pollution in China. Fine particles (PM2.5, diameter < 2.5μm) are much more hazardous than larger particles with respect to mortality, cardiovascular and respiratory endpoints, and PM2.5 is considered to be the best indicator of the level of health risk resulting from air pollution by the WHO.[6] However, real time PM2.5 information was not included in the calculation of API. As they had little or no information about this important determinant of air quality, Chinese people used to regularly understate the level of

[4] Roback (1982) points out that hedonic prices of different amenities can be used as weights in the calculation of a city quality of life (QOL) ranking. Our analysis suggests that information constraints will also lead to a wedge between the estimated and true QOL rankings.

[5] API is defined as the maximum value of three pollutant indexes, including SO2, NO2, and PM10.

[6] See WHO reports: http://www.who.int/mediacentre/news/releases/2014/air-quality/en/

atmospheric contamination, and would often mistake low visibility on polluted days caused by severe PM2.5 pollution for fog rather than smog (Barwick et al., 2019).

In 2013, China launched a nation-wide, real-time air quality monitoring and disclosure program. The program published a real-time Air Quality Index (AQI) and information about PM2.5,[7] incorporating PM2.5 concentration into air quality measures for the first time in China. To address the issue with the previous reporting system, the program established a monitoring network delivering real-time air quality data to the central government, effectively enforcing pollution monitoring and significantly improving data accuracy.

The nation-wide program was conducted in three waves on the basis of pre-determined conditions, including the location (e.g. Yangtze River Delta, Pearl River Delta), the city tier (e.g. provincial-level cities), and pre-determined designations (e.g. national environmental protection exemplary cities designated prior to the program).[8] In the first wave, 74 major Chinese cities released real-time data on PM2.5 and other air pollutants by December, 2012. The second wave added 116 cities by October 31, 2013, and the remaining 177 cities joined the program by November 20, 2014 in the final wave. The information on real time PM2.5 concentration has been available to all Chinese cities by 2015.

The program published both hourly and daily PM2.5 data in real time on the website of the Ministry of Environmental Protection of China (MEP), and mass media was encouraged to disseminate the data. The sudden disclosure and dissemination of real time air quality data dramatically improved the public access to local pollution information and dramatically increased public awareness of the health costs of pollution exposure. The unexpected data disclosure strongly affected the avoidance behavior of Chinese citizens. As illustrated in Figure A1, the transaction of indoor air filtration increases sharply in response to the information shock.

Khanna, Liang, Mobarak and Song (2021) document that people tend to sort from polluted to clean cities in China, and the PM2.5 data disclosure significantly enhanced the migration response to air pollution. The strength of the Rosen-Roback hedonic framework lies in the use of income and housing price differentials across space driven by amenity-induced sorting. Therefore, the information disclosure should impact the hedonic price of an air quality improvement measured

[7] Since 2013, China started to release real time data on six major air pollutants, including PM2.5, P10, O3, CO, NO2 and SO2. The AQI is an overall index of these major air pollutants.

[8] See *the Implementation Plan for the First Phase Monitoring of the New Air Quality Standards*, *the Implementation Plan for the Second Phase Monitoring of the New Air Quality Standards* and *the Implementation Plan for the Third Phase Monitoring of the New Air Quality Standards* released by the Ministry of Environmental Protection of China (MEP).

by the Rosen-Roback framework. Figure 1a shows that there was no detectable relationship between real income (i.e., nominal income adjusted by housing costs, defined formally in Section 5) and the amount of PM2.5 across space when PM2.5 data were not available to the public. However, as illustrated in Figure 1b, we see a clear positive association between PM2.5 concentration and real income after the publication of real-time PM2.5 data. The positive compensating differential for PM2.5 exposure after the data disclosure is in accordance with the spirit of Rosen-Roback theory.

4. Data

4.1 Air Quality Data

We use satellite-derived PM2.5 data, covering periods before and after China's disclosure of official PM2.5 data. City-level annual PM2.5 concentrations are measured using the Global Annual PM2.5 Grids derived from satellite data by Van Donkelaar et al. (2016).[9] The raster grids of this ground calibrated PM2.5 data have a high grid cell resolution of 0.01 degree. This yields a comprehensive and reliable measurement of air quality for a wide range of cities in China, covering all the prefecture, sub-provincial and provincial cities. Figure A2 shows that our satellite-based measure of PM2.5 is consistent with the official measure of PM2.5 released by the MEP in China. The correlation between satellite PM2.5 data and monitor-based PM2.5 data in China is up to 0.8 (Freeman, Liang, Song Timmins, 2020).

4.2 Income and Housing Price Data

Income and housing price data come from China Labor-force Dynamics Survey (CLDS). CLDS is a national social survey, covering information of around 21,000 individuals in about 14,000 households across 29 provinces of China. A probability-proportional-to-size sampling (PPS) based on population size and administrative units is adopted to ensure the national representatives of the survey. As a result, the distribution of sample size across cities in CLDS is consistent with the geographic distribution of population in China.

The data not only record very detailed information of housing prices and housing conditions, but also contain a wide range of demographic and social economic characteristics of individuals,

[9] Van Donkelaar et al. (2016) estimate ground-level PM2.5 by combining Aerosol Optical Depth (AOD) retrievals from the NASA MODIS, MISR, and SeaWiFS, which are subsequently calibrated to global ground-based observations of PM2.5 using Geographically Weighted Regression (GWR).

including age, gender, education level, residential city, employment status, *hukou* type (rural/urban), and income. We restrict our attention to the working population with positive income. On average, wage income accounts for 97.3% of individual income in our sample. As housing attributes and ownership may confound the location-specific housing price measure, we follow Bayer et al. (2007, 2009) to use micro housing information to compute the city-specific housing price, regardless of housing attributes and ownership[10].

CLDS is conducted in three waves in 2012, 2014 and 2016, recording information about income and housing price one-year prior to the survey. Thus, the data cover periods both before and after the publication of official PM2.5 data in China. We combine three waves of the survey, and construct individual-level pool cross-section data across 2011, 2013 and 2015 for our empirical analysis.

4.3 Inputs into Instrumental Variables and City Controls

We collect information on large-scale (capacity>1.5 million KW) thermal power plants, and their coal consumption from China's Electric Power Yearbooks and Energy Statistical Yearbooks. We supplement this with information on the establishment year of plants, the angle between their locations and annual dominant wind direction of each city, and the distance to each city.

We obtain city characteristics, involving GDP per capita, government expenditure and other amenities from the City Statistical Yearbooks and China Urban Construction Statistical Yearbook. Weather condition data come from China Meteorological Data Service Center. We collect monthly data on weather amenities and convert into corresponding yearly measure.

5. Empirical Specification

In this section, we empirically test the predictions of our variant of Rosen-Roback theory incorporating the role of information. We rearrange equation (4) and (7) in section 2, and the MWTP for improvement to a particular amenity X_j can be estimated as:

$$MWTP_X = \frac{\partial Log\left(I_{i,j}/\rho_j^{\theta}\right)}{\partial X_j} \times I_{i,j} \tag{9}$$

$I_{i,j}/\rho_j^{\theta}$ measures real income – i.e., housing price adjusted income[11]. The parameter θ measures

[10] See Appendix B for the detailed procedure to calculate city-specific housing price.

[11] Tombe and Zhu (2019) and Khanna, Liang, Mobarak and Song (2021) use the same specification to measure real income in

the fraction of housing expenditure in income and captures the relevance of housing costs in the deflation of income in this expression. Thus, our real income measure represents the spatial variation in both nominal income and housing prices, and we use the real income gradient with respect to the level of the amenity to calculate the corresponding hedonic price.

Our theoretical analysis documents that the MWTP estimates differ in scenarios with and without the public access to amenity information, so we leverage the natural experiment of unexpected PM2.5 data disclosure in China to estimate changes in the hedonic price of air quality driven by the information shock. Our empirical specification is as follows:

$$Log\left(Real\ Income_{ij,t}\right) = \beta_0 + \beta_1 PM2.5_{jt} \times Disclosure_{jt} + \beta_2 PM2.5_{jt} + \beta_3 Disclosure_{jt} \quad (10)$$
$$\beta_4 Z_{ij} + \xi_j + \eta_{Region,t} + \delta_{\ Tier,t} + \varepsilon_{ij,t}$$

where $Real\ Income_{ij,t}$ is the real income received by individual i in city j and year t. $Disclosure_{jt}$ is an indicator for whether or not real time PM2.5 data have been published in city j and year t, and $PM2.5_{jt}$ is the amount of PM2.5 concentration. Our primary independent variable of interest is the interaction of the amount of PM2.5 to the information disclosure indicator, which represents the wedge between the revealed hedonic price and the implicit value of clean air under information constraints. Our modified Rosen-Roback model documents that the wedge depends on people's perception bias under incomplete information, and the parameter λ in the model measures the direction and the magnitude of that bias. We can directly link equation (10) to our model, and $\lambda = \beta_0/(\beta_0 + \beta_1)$. If the perception parameter λ<1, people understate the level of air pollution before the data disclosure.

The schedule of staggered publication of PM2.5 data depends on tier and location of cities and other pre-determined conditions, so we control for city fixed effects ξ_j, and city tier-by-year fixed effects $\delta_{\ Tier,t}$. We also include region-by-year fixed effects $\eta_{Region,t}$ to account for spatial differentiated environmental and economic policies in China[12]. We use individual-level pooled cross-sectional data in 2011, 2013 and 2015 to estimate equation (10). Z_{ij} are controls, including individual demographic characteristics and Dahl correction terms to address the potential issue of

China.

[12] There are seven macro-regions in China: East China, North China, Central China, South China, Southwest China, Northwest China and Northeast China.

Roy sorting bias (Roy, 1951; Dahl, 2002)[13].

Our specification explores the exogenous changes in the public access to pollution data. The central government determined the sequence of staggered data publication according to the tier and the location of cities, as well as other pre-determined designations. Since these conditions are pre-determined prior to the program, controlling for city-fixed effects along with city tier-by-year fixed effects allows us to account for unobservables correlated with the staggered sequence of the program. In Table A2, we test whether lagged city attributes can predict the sequence of information disclosure across cities controlling for these fixed effects. We fail to find any meaningful associations between the unexpected information release and lagged city attributes, like population, GDP, energy consumption, pollutants emission and industrial structure.

Pollution is likely to be associated with local economic activity, so naïve OLS estimates may be biased as a result. To deal with the endogeneity concern, we use two different strategies to isolate the causal effect of pollution—an instrumental variable based on how wind direction interacts with the placement of distant thermal power plants (as in Freeman et al., 2019), and a regression discontinuity around the Huai river (as in Chen et al., 2013).

5.1 Instrument #1: Wind Direction and Coal-Fired Power Plants

Our first source of plausibly exogenous variation in pollution is based on an insight from Freeman, Liang, Song, and Timmins (2019). We quantify the extent to which distant large-scaled thermal power plants are located upwind of a given city. The instrument value is penalized if plants are not located directly upwind of the city and if it is farther away, using our first-stage equation:

$$PM2.5_{jt} = \gamma_0 + \gamma_1 \sum_p^P \left(\frac{1}{\alpha_{pt}+1}\right)\left(\frac{1}{dist_{pj}}\right) C_{pt} + \xi_j + \eta_{Region,t} + \delta_{\ Tier,t} + \mu_{jt} \tag{11}$$

where α_{pt} is the angle between the annual dominant wind direction of city j and the plant p, and changes across years, $dist_{pj}$ denotes the distance from the plant p to city j , C_{pt} is the coal consumption of plant p in year t . We only use large-scale thermal power plants that are located

[13]A potential issue with the wage hedonic estimation is the Roy sorting bias. Roy sorting refers to the problem that individuals respond to idiosyncratic wage draws and are likely to move to a location where that wage draw is good. For example, individuals from a particular region could earn unusual high wages in a given place, because their personal abilities have unusual comparative advantages specific for working in this place. Thus, other people who looks like these individuals cannot earn same wages if they move to the place. We follow the semi-parametric approach proposed by Dahl (2002) to address the Roy sorting bias by adding correction terms. See Appendix C for detailed discussions on Roy sorting and the approach of Dahl (2002). Addressing the Roy sorting bias also helps us to account for mobility costs faced by migrants; mobility costs may affect wages earned by migrants.

more than 50 km from a city, but within a 300km from the city. Figure A3a describes the intuition of the instrument. The underlying variation comes from how wind patterns blow air pollutants from distant thermal power plants to cities. Both yearly wind direction and the number of power plants change overtime. Figure A4 shows that the first stage relationship between the wind direction IV and local air pollution is strong, and Table A4 reports that the F-statistics are all greater than 10 across different specifications.

We expect that our instrumental variable is orthogonal to local economic activities. First, wind direction is determined by nature, and it is exogenous to local economic activities. Second, those large-scale thermal power plants supply electricity to vast areas of China, including many remote regions; many even do not supply electricity at all to their nearby cities, but rather to many remote provinces. Further, the setup of large-scale power plants and the allocation of electricity supply from them is determined by the central government – it is difficult for local governments to exert influence on the siting of large-scale plants and the electricity supply from them.[14] Finally, the spillover from distant power plants on local economic activity is extremely small, but the pollutants emitted from power plants located upwind severely contaminate the local air.

We examine potential concerns with this instrument in Section 6.2. For instance, the location and the coal consumption of power plants tend to be correlated with economic conditions of nearby cities, and so we demonstrate robustness to excluding power plants in various distances from cities. Our falsification tests further indicate that baseline economic characteristics, population and electricity demand do not predict the future placement of plants. Both the public and the government's concern over environmental quality are relatively recent, and we show robustness to using only old power plants, such as those built before the establishment of the MEP. Among other sensitivity tests, the IV results are strongly robust to excluding populated and politically important cities and coal-producing provinces, as well as adding additional controls for electricity consumption and economic conditions.

5.2 Instrument #2: The Huai River Regression Discontinuity

Our second identification strategy relies on an important spatial-differentiated public heating policy in China. Since the 1950s, China established a coal-based free heating system to cities north of the Huai River. This policy had long lasting effects, as even today the heating supplies

[14] In 2015, large-scale (capacity>1.5 million KW) thermal power plants only accounts for 2.59% of coal-fired plants in China.

differ largely between cities on different sides of the river. The large amount of coal burning used for centralized heating supply has substantially driven up the level of air pollution in northern cities. As shown in Figure A5, there is a sharp increase in PM2.5 concentration just north of the River border. Chen et al. (2013) use this spatial policy discontinuity to study the impacts of air pollution on life expectancy by comparing cities straddling cross the Huai River.

Driven by *the Action Plan on Air Pollution Prevention* released by the State Council of China, cities in the north of China started to gradually switch from coal to natural gas for winter heating beginning in 2013. Thus, our second plausibly exogenous source of variation in pollution comes from the interaction of the long-lasting public heating policy to the recent fuel switching policy. The interplay of two polices causes the difference in air quality between the north and the south to change over time. We will discuss the Huai River regression discontinuity in detail in section 6.2.4.

5.3 Identification on how Pollution Effects Shift with the Information Shock

We focus on how the real income-pollution gradient changes in response to the unexpected publication of PM2.5 data. Our primary variable of interest is the interaction of an information publication indicator with the level of PM2.5, and we control for the two independent components of interaction term simultaneously. Including the information dummy accounts for any remaining unobservables that may be correlated with the program, and adding the level of pollution accounts for unknown confounders that may not be fully accounted for by our instruments. The validity of our identification strategy is therefore based on the assumption that there are no remaining unobservables that are systematically correlated with the instruments. In other words, the nature of the potential endogeneity concern of our instruments does not change before and after the information shock. To examine this assumption, we perform balance tests in Table A3. There are no systematic differences in both the level of air pollution and the value of our wind direction IV before and after the unexpected disclosure of pollution data. Additionally, the information disclosure does not have any meaningful effect on a wide range of city characteristics, including other amenities, government environmental protection expenditure, industrial structure, emission discharges, and among others.

6. Empirical Results

6.1 Baseline Results

Table 1 presents the baseline results of the shift in the real income gradient with respect to PM2.5 concentration driven by the unexpected disclosure of real-time PM2.5 information. Panel A shows the OLS estimates. Both the amount of PM2.5 and its interaction with the information disclosure indicator are statistically insignificant. Local PM2.5 concentration tends to be correlated with industrial production, population and other confounding factors, which may affect earnings and housing prices. Thus, the endogeneity problem biases the naïve OLS estimates, resulting in both statistical and economic insignificance of the coefficient estimates.

In Panel B, we deal with the endogeneity concern using the instrumental variable based on the interaction between wind direction, the location and the coal consumption of distant coal fired plants. In column (1), we control for both city fixed effects and region-by-year fixed effects to account for city-level predetermined characteristics (prior to the information program and may be associated with its implementation) and spatially differentiated public policies in China. Since the sequence of the information policy rollout depends on the tier of cities, we further add the city tier-by-year fixed effects in column (2). Column (2) presents the results of our baseline specification of Equation (10), and we impose a more stringent restriction by controlling for a triple interaction between region-, city tier- and year- fixed effects in column (3). Our IV estimates are robust to different specifications. Across all these specifications, the coefficients on the interaction of PM2.5 concentration and the PM2.5 data disclosure dummy are positive and statistically different from zero. The results indicate that the release of official PM2.5 data raises public awareness of air pollution and significantly increases the compensating differential required for exposure to PM2.5 pollution. The coefficient on the level of PM2.5 alone is positive but insignificant, indicating no significant association between PM2.5 concentration and real income before the public had access to local PM2.5 information. We further include weather amenities in columns (4)-(6) to rule out the potential for endogeneity due to the correlation between pollution and weather conditions. Our parameter estimates are quantitively and qualitatively similar.

The IV estimates of the interaction term range from 0.0084 to 0.0087 and the estimates of the information disclosure indictor range from -0.34 to -0.35. Since the median level of PM2.5 is around 40 in China, combining the two parameter estimates implies that the release of pollution information raises real income in highly polluted cities to compensate for the adverse effects of pollution, but reduces real earnings in less-polluted cities.

The empirical results are consistent with the predictions of the variant of Rosen-Roback theory sketched in Section 2. Incomplete information biases the hedonic price estimates and creates a wedge between the revealed hedonic price and the true implicit value of amenities. As demonstrated by our model with information constraints, the direction of the estimation bias depends on the direction of people's perceptual bias. The corresponding perception bias parameter in the model $\lambda = \beta_0/(\beta_0 + \beta_1) = 0.5$, and the coefficient on the interaction of data disclosure indicator and the level of PM2.5-- i.e. β_1 -- is significantly positive. Thus, people tend to underestimate the level of air pollution before the disclosure of pollution information. This is in accordance with the finding of Barwick et al. (2019) that Chinese citizens used to regularly understate the severity of atmospheric contamination, and would mistake low visibility on polluted days caused by severe PM2.5 pollution for fog rather than smog (Barwick et al., 2019). As a consequence of the downward perceptual bias, there is a downward estimation bias in the hedonic price of clean air due to incomplete information.

Referring to our baseline results in column (2), a median Chinese citizen is willing to pay 169 Chinese Yuan for a 1 $\mu g/m^3$ reduction in PM2.5 concentration under incomplete information, and 337 Chinese Yuan for the same marginal reduction in PM2.5 concentration under complete information. The shift in the MWTP estimates driven by the data disclosure accounts for approximately 0.84% of the median income in China.

6.2 Sensitivity Analysis

Our baseline empirical results confirm the conclusions of our variant of Rosen-Roback model with information availability. In this section, we perform a wide range of meaningful robustness checks. We explore threats to identification, different source of variation, alternative model specifications and data sources.

6.2.1 Endogeneity Concerns over the Wind-Direction IV

In our baseline empirical specification, we employ the instrumental variable based on the interplay between wind direction, the distance to cities and the coal consumption of distant large-scale coal-fired plants to address the endogeneity issue of air pollution. A major issue with this IV is the potential endogenous placement of power plants. Policy makers may take the three components of the IV into account when placing thermal power plants, and they may tend to protect the air quality of certain types of cities. Thus, address the concern that this IV is correlated with unobservables associated with nearby cities. In Table A6, we further exclude the power plants within 80km, 130km and 180km around the city, respectively. Our empirical pattern is pretty robust to excluding nearby coal-fired plants.

Even though policy makers may not have used the same criteria – the interaction between wind direction, the distance to cities and the coal consumption – to site power plants in the past, they may have paid more attention to the resulting environmental costs when locating power plants in more recent years. Both the public and the government have recently started to attach more importance to environmental protection in China; an important signal of this was the establishment of the Ministry of Environmental Protection (MEP) in 2008. We thus exclude power plants built within the most recent five to eight years, respectively, in Table A7. Indeed, when we use power plants built more than eight years ago, we rely only on plants built prior to the establishment of the MEP in China. This empirical strategy is more conservative, because cities with newly built plants are assigned to the 'control group'. However, we still see a similar empirical pattern. The disclosure of pollution information significantly raises the real income-pollution gradient in the Rosen-Roback framework, and hedonic price of avoiding PM2.5 exposure increases around 152-181 Chinese Yuan. Therefore, we rule out the concern that our IV results are confounded by the recent environmental concern and policies in China.

We may expect that thermal power plants are more likely to be located in coal-producing regions in China. The concentration of coal production may affect regional industrial structure, raising concerns over other unobserved correlations with individual income. Shanxi is the largest coal producing province in China, we therefore drop Shanxi province in Table A8. Excluding coal producing region does little to influence the association between pollution information and the estimated hedonic price in our Rosen-Roback Framework.

Another concern with this IV strategy is that the siting and coal consumption of power plants may be driven by the electricity demand of nearby cities, even though large-scale plants supply electricity to vast areas in China, including many remote provinces (Freeman, Liang, Song, and Timmins, 2019). To account for the potential role played by electricity demand, we control for city-level industrial electricity consumption (in Panel A), and then total electricity consumption (in Panel B) in Table A9. Controlling for electricity demand barely changes our empirical pattern. The MWTP to avoid exposure to PM2.5 pollution increases from 156 to 172 Chinese Yuan in response to the information shock, which is quite similar with our baseline estimates.

To further test whether policy makers favor populated and politically important cities, and locate power plants in a way that takes the interplay between wind direction, the location and the coal consumption of power plants into account. We examine whether baseline city attributes can predict newly built plants. In Table A10, we look at the predictive effects of city characteristics in 2004 on the ratio of upwind power plants built after 2005, as well as the IV constructed only using plants built after 2005. Baseline city characteristics, including GDP, population and electricity demand, have nothing to do with the location and coal consumption of newly built coal-fired plants.

6.2.2 An Alternative Version of the Wind-Direction IV

In this section, we construct an alternative version of the wind-direction IV that the coal consumption of power plants located upwind of a given city as the IV, controlling for the aggregate coal consumption of all other power plants located not in the upwind area but within the same distance radius from the city. The empirical specification of first stage and second stage IV estimation is as follows:

$$PM2.5_{jt} = \gamma_0 + \gamma_1 \sum_{p,Upwind}^{P} \left(\frac{1}{dist_{pj}}\right) C_{Pt} + \gamma_2 \sum_{p,Non-upwind}^{P} \left(\frac{1}{dist_{pj}}\right) C_{pt} \qquad (12)$$

$$+\xi_j + \eta_{Region,t} + \delta_{\ Tier,t} + \mu_{j,t}$$

$$Log\left(Real\ Income_{ij,t}\right) = \alpha + \beta_1 PM2.5_{jt} + \beta_2 \sum_{p,Non-upwind}^{P} \left(\frac{1}{dist_{pj}}\right) C_{pt} \qquad (13)$$

$$+\xi_j + \eta_{Region,t} + \delta_{\ Tier,t} + \varepsilon_{ij,t}$$

where $\sum_{p,upwind}^{P} \left(\frac{1}{dist_{pj}}\right) C_{pt}$ represents the total coal consumption of plants located in the upwind area of city j, and the value is penalized if those plants are further away. As illustrated in Figure A3b, the upwind area is a section of a circular buffer drawn at a distance of 50-300km from city j, and the angle between the left/right side of the section and the wind direction of the city is 45 degrees. $\sum_{p,Non-upwind}^{P} \left(\frac{1}{dist_{pj}}\right) C_{pt}$ denotes the total coal consumption of plants located within the same distance (50-300km) from city j but not in the upwind area, and the value is penalized in the same way based on distance. In Figure A3b, the shaded dark grey area is the upwind area of city j. We exclude the upwind area around the city, and the light grey area is defined as the non-upwind area around city j.

The variation of this alternative wind-direction IV comes only from how wind direction allocates coal consumption between the upwind versus the non-upwind areas around a particular city. Although we consider the distance to the city and the coal consumption of upwind plants in the IV construction in Equation (12), we account for the distance and the coal consumption of those 'controlled' plants located in the non-upwind area around the city. Therefore, this IV strategy allows us to account for any remaining unobservables that may be correlated with total coal consumption and distance to cities.

Table A11 reports the first-stage estimates. Because we only leverage the distribution of coal consumption between the upwind area versus other areas in this IV, the first stage relationship is not so strong as that of our baseline wind direction IV, but the F-statistics are still close to 10 across different specifications.

Table A12 presents the corresponding second stage results. The empirical pattern is quantitively and qualitatively similar to our baseline results. The coefficients on our primary variable of interest – the interaction between PM2.5 concentration and the data disclosure indicator – is precisely estimated and significantly positive. The results are also insensitive to excluding power plants within various distance bins of cities. Therefore, our alternative wind direction IV yields a similar empirical pattern that incomplete information leads to a wedge between the hedonic price and implicit value of clean air, and release and dissemination of air quality data kills such wedge.

The only potential concern with this alternative IV is the endogeneity of wind direction. Policy makers may locate coal-fired plants such that pollutants do not travel to politically important or heavily populated cities. If that were the case, we should see fewer thermal plants placed upwind of these influential cities. Table A13 reports the number of large-scale coal-fired plants sited upwind of five largest cities as well as the total amount of their coal consumption in 2015. Beijing and Tianjin are the most populated and politically important cities in Northern China, but they experience severe atmospheric contamination. However, we see three large-scale coal-fired plants located upwind of Beijing and Tianjin, respectively. The total consumption of plants located upwind of Beijing is up to 195 million tons, and is 150 million tons for those located upwind of Tianjin. In contrast, an average Chinese city only has two upwind large-scale plants with the total coal consumption of 79 million tons, which are far below the values in Beijing and Tianjin. We next compute the fraction of the upwind plants in the total number of large coal-fired plants in the second column. The ratio is 38% for both Beijing and Tianjin, slightly higher than the corresponding national mean. Overall, the summary statistics of the coal consumption and the location of power plants indicate that policy makers do not intentionally place the coal-fired plants away from these populated or politically important cities. This is in line with the falsification test showing that baseline city population and economic conditions do not predict the future placement of plants in section 6.2.1.

6.2.3 Additional Controls and Alternative Samples

In this section, we first introduce various sets of covariates that might confound the relationship between public access to PM2.5 information and the compensating differential for PM2.5 exposure. We then examine whether our empirical pattern is driven by big cities and highly polluted cities.

The strength of the Rosen-Roback theory lies in that regional differences in income and housing costs reflect the implicit values of amenities. Thus, spatial differentiated economic conditions may confound the estimated compensating differential for air pollution exposure. As reported in Panel A of Table A14, our estimates display similar patterns as before if we include GDP per capita, population and industrial structure as covariates. Therefore, the changes in the MWTP estimates (revealed by the Rosen-Roback Framework) in response to the information dissemination is not driven by economic conditions.

Air quality is only one of the important amenities determining location-specific livability, so based on the spirit of the Rosen Roback theory, regional differences in real income may pick up the effects of other amenities. To allay this concern, we control for other local amenities affecting the welfare of residents in Panel B of Table A14, including the number of doctors, the number of library books, and the area of green coverage. Adding additional amenities does not affect our results meaningfully, indicating that our IV strategy does a good job to isolate the effects of air quality.

We expect that fine particulate matter concentration may correlate with local industrial emissions. To account for the confounding effects of local pollutant discharges, we control for industrial water emission, industrial SO_2 emission and industrial dust emission in Panel C of Table A14. Our estimated empirical pattern hardly changes, and the effects of information disclosure on the estimated MWTP from the Rosen-Roback framework is quite similar to our baseline estimates.

In developing countries like China, industrialization, severe pollution and economic opportunities are highly concentrated in several big cities, which may confound the relationship between pollution and spatial income differences in the Rosen-Roback Framework. To examine whether our hedonic estimates are driven by these cities, we exclude one influential city at a time in Table A15. Our results are quantitively and qualitatively similar.

6.24 The Huai River Regression Discontinuity

In this section, we use a different identification strategy based on the spatially-differentiated winter heating system in China. Beginning in the 1950s, China has provided a centralized coal-based heating system to cities north of the Huai River. This policy has had long lasting effects, as even today, cities north of the river boundary receive centralized heating supply from the government every winter, whereas cities in the south do not (Ito and Zhang, 2020). The heating system relies on coal burning in water boilers, releasing a large amount of particulate matter and leading to a higher level of air pollution in the north.

In response to *the Action Plan on Air Pollution Prevention*, China started to gradually switch coal to natural gas for winter heating after 2013, which has attenuated the effects of the centralized heating on air quality deterioration in northern cities. Thus, we leverage the variation in pollution coming from the interplay of the long-lasting public heating policy to the recent fuel policy. Driven by the interaction of the two polices, the difference in air quality between the north and the south would change over time. Thus, the first-stage equation is given by:

$$PM2.5_{jt} = \psi_0 + \psi_1 North_j \times Year\ FE_t + \psi_2 R_j \times North_j \times Year\ FE_t \\ + \xi_j + \eta_{Region,t} + \delta_{\ Tier,t} + \nu_{jt} \tag{14}$$

$North_j$ is an indicator variable for whether city j is located north of the Huai River boundary, and the running variable R_j is the distance of city j to the river border. We use the interaction of north of Huai River indicator and year-fixed effects to leverage the overtime variation in the discontinuous shift in fine particle concentration at the river border (R_j =0). The strength of this specification lies in that it allows us to control for city-fixed effects to account for city-level unobsevables that may confound the effects of information disclosure.

We follow Ito and Zhang (2020) to select 400 miles as the bandwidth, and use a local linear control for the running variable. We interact the linear control with year-fixed effects to allow for changes in the effects of distance to the river border overtime. Table A5 presents the first-stage results of our spatial regression discontinuity, with F-statistics greater than 10 across various specifications. The north-south air quality gap is significantly larger in 2011 and 2013, in comparison with that in 2016 when China has largely switched to natural gas for winter heating in northern cities. Table A16 presents the second stage RD estimates. Once again, we find a similar empirical pattern suggesting a significant increase in the MWTP for reducing PM2.5 concentration in response to the pollution information rollout. We control for the interaction of the longitude decile to year-specific fixed effects in the last three columns to allay the concern that unobservables in the west-east dimension may confound our RD estimates. However, accounting for these unobservables barely changes our estimates. As reported in Table A17 and A18, our spatial RD estimates are robust to the selection of bandwidth and control function for the running variable.

6.2.5 Placebo Tests

To further examine the causal relationship between the unexpected information disclosure and the shift in the hedonic price of avoiding PM2.5 exposure, we conduct various sets of placebo tests on our IV identification as well as the treatment effects of the unexpected information disclosure.

Our first IV is based on the plausibly exogenous variation driven by how wind direction interacts with the location and the coal consumption of distant coal-fired plants. Section 6.2.2 documents that we can consistently find similar results, if we fix the total coal consumption of power plants and only leverage how wind direction allocates coal consumption between the upwind area and other areas within the same distance radius. This indicates that wind direction is the dominant factor driving our empirical pattern (rather than the distance to cities and the total coal consumption of plants). To further test the validity of this approach, we perform a placebo test by adding 180 degrees, to the wind direction angle, while holding all other factors in the IV construction constant. As reported in Table A19, the falsified instrument can hardly identify the effects of air pollution exposure, confirming that wind direction is the driving factor determining our baseline IV results.

Our second source of plausibly exogeneous variation lies in the interplay of the centralized heating policy in cities north of the Huai River and the recent energy policy of switching coal to natural gas for winter heating. The cutoff for our RD design is the Huai River boundary, we thus move the river border in a parallel fashion by 5 degrees in Table A 20. The placebo spatial RD is less likely to predict the effects of our primary variable of interest.

We next turn our attention to the treatment effects of the data rollout in China. The real time PM2.5 data were released in three waves, and the sequence of data rollout determine the treatment status of cities. We assign a "placebo" treatment status to each city by randomly allocating cities to each of the three waves. Table A21 shows that the "placebo" treatment of data disclosure does not have any meaningful effect on the hedonic price of reducing PM2.5 exposure. In Table A22, we conduct additional placebo tests by arbitrarily delay the time of data release by one year and two years, respectively. The placebo data disclosure is unable to pick up the significant change in the MWTP for clean air induced by the improved access to information, predicted by our variant of Rosen-Roback theory sketched in Section 2.

7. Conclusion

Our paper highlights the consequences of restricted access to information on non-market valuation. The core of the Rosen-Roback theory lies in that the hedonic price of an amenity measured by spatial differences in income and housing prices is equivalent to its implicit value. Our theoretical analysis demonstrates that information failure would differentiate the estimated hedonic price from the true implicit price of the amenity, undermining the core of the Rosen-Roback framework. Changes in the hedonic price of avoiding PM2.5 exposure, driven by the natural experiment of unexpected PM2.5 information disclosure in China, consolidate the predictions of our theoretical analysis. Since the public has limited or no access to real time environmental information in many developing countries, our theoretical and empirical evidence directly speaks to behavioral paradox between the extreme high economic and health burden of pollution and people's unreasonable low MWTP for improvement to environmental quality in developing countries (Greenstone and Jack, 2015).

Our work has important implications for cost-benefit analysis of environmental policies. The goal of Pigouvian policies is to force polluters to internalize the social costs of polluting activities in the absence of a market (Pigou, 1912). Information constraints would bias the estimates of the social costs of pollution, rendering it much harder for policy makers to design optimal environmental policies.

Our analysis documents that estimation bias in hedonic valuation is driven by perception biases that arise due to limited information. With the increased availability of global satellite-based data on various dimensions of environmental quality, future researchers can continue to measure the difference between levels of environmental quality perceived by the public and measured by satellite data and develop approaches to recover estimates of the true willingness to pay under incomplete information.

References

Albouy, D.Y., 2008. *Are Big Cities Really Bad Places to Live?: Improving Quality-of-life Estimates Across Cities*. *NBER Working Paper*, (w14472).

Aldy, J.E. and Viscusi, W.K., 2008. Adjusting the value of a statistical life for age and cohort effects. *The Review of Economics and Statistics*, 90(3), pp.573-581.

Barwick, P.J., Li, S., Lin, L. and Zou, E., 2019. From Fog to Smog: The Value of Pollution Information. *NBER Working Paper*, (w26541).

Bayer, P., Ferreira, F. and McMillan, R., 2007. A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, 115(4), pp.588-638.

Bayer, P., Keohane, N. and Timmins, C., 2009. Migration and Hedonic Valuation: The Case of Air Quality. *Journal of Environmental Economics and Management*, *58*(1), pp.1-14.

Blomquist, G.C., Berger, M.C. and Hoehn, J.P., 1988. New Estimates of Quality of Life in Urban Areas. *The American Economic Review*, pp.89-107.

Bryan, G. and Morten, M., 2019. The aggregate productivity effects of internal migration: Evidence from Indonesia. *Journal of Political Economy*, 127(5), pp.2229-2268.

Chen, Y., Ebenstein, A., Greenstone, M. and Li, H., 2013. Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences*, 110(32), pp.12936-12941.

Chen, Y., Jin, G.Z., Kumar, N. and Shi, G., 2012. Gaming in Air Pollution Data? Lessons from China. *The BE Journal of Economic Analysis & Policy*, 13(3).

Dahl, G.B., 2002. Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica*, 70(6), pp.2367-2420.

Freeman, R., Liang, W., Song, R. and Timmins, C., 2019. Willingness to Pay for Clean Air in China. *Journal of Environmental Economics and Management*, 94, pp.188-216.

Ghanem, D. and Zhang, J., 2014. 'Effortless Perfection:' Do Chinese Cities Manipulate Air Pollution Data?. *Journal of Environmental Economics and Management*, 68(2), pp.203-225.

Greenstone, M. and Jack, B.K., 2015. Envirodevonomics: A Research Agenda for An Emerging Field. *Journal of Economic Literature*, 53(1), pp.5-42.

Hoch, I. and Drake, J., 1974. Wages, Climate, and the Quality of Life. *Journal of Environmental Economics and Management*, 1(4), pp.268-295.

Hersch, J., 1998. Compensating Differentials for Gender-specific Job Injury Risks. *The American Economic Review,* 88(3), pp.598-607.

Ito, K. and Zhang, S., 2020. Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Journal of Political Economy*, 128(5), pp.1627-1672.

Kahn, M.E., 1995. A Revealed Preference Approach to Ranking City Quality of Life. *Journal of Urban Economics*, 38(2), pp.221-235.

Khanna, G., Liang, W., Mobarak, A.M. and Song, R., 2021. The Productivity Consequences of Pollution-induced Migration in China. *NBER Working Paper*, (w28401).

New York Times, 2015, December. Companies in South China See Opportunity in Beijing's Smog. The New York Times Company. https://www.nytimes.com/2015/12/23/world/asia/beijing-air-pollution-china-smog.html.

Pigou, A.C., 1912. *Wealth and Welfare*. Macmillan and Company, limited.

Roback, J., 1982. Wages, Rents, and the Quality of Life. *Journal of political Economy*, 90(6), pp.1257-1278.

Rosen, S., 1979, Wages-Based Indexes of Urban Quality of Life, In Mieszkowski P. and Straszheim M.(eds.), Current Issues in Urban Economics, Baltimore, *John Hopkins University Press*, 74-104.

Roy, A.D., 1951. Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2), pp.135-146.

Tombe, T. and Zhu, X., 2019. Trade, Migration, and Productivity: A Quantitative Analysis of China. *American Economic Review*, 109(5), pp.1843-72.

Wang, Z. and Zhang, J., 2021. The Value of Information Disclosure: Evidence from Mask Consumption in China. *Duke Kunshan University Working Paper.*

Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., Lyapustin, A., Sayer, A.M. and Winker, D.M., 2016. Global Estimates of Fine Particulate Matter using a Combined Geophysical-statistical Method with Information from Satellites, Models, and Monitors. *Environmental science & technology*, 50(7), pp.3762-3772.

Viscusi, W.K., 1993. The value of risks to life and health. *Journal of economic literature*, 31(4), pp.1912-1946.

Figure 1a: PM2.5 and Real Income *before* the Information Disclosure

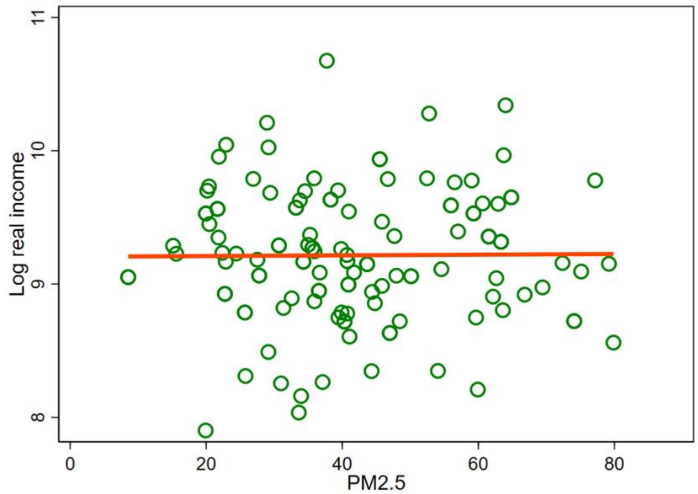
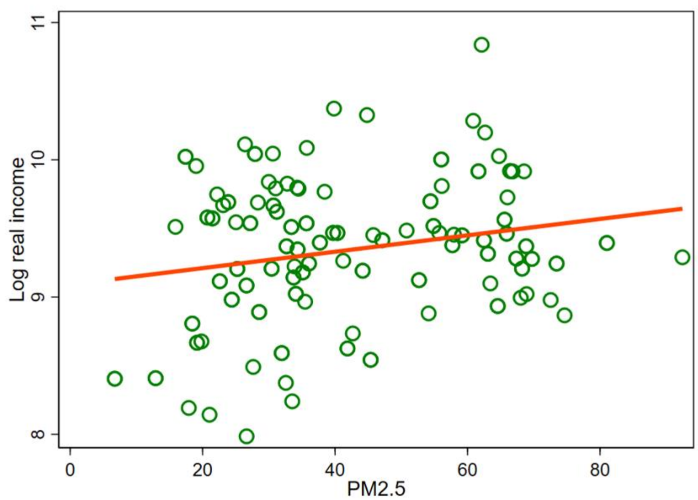


Figure 1b: PM2.5 and Real Income *after* the Information Disclosure



Notes: Real Income is measured by I/ρ^{θ}. I and ρ are nominal income and housing price, respectively, and θ is the share of housing expenditure in income. Cities are grouped into one hundred groups according to the quantile of PM2.5 concentration. The y-axis denotes the log of the mean value of real income in each quantile, and x-axis denotes the mean value of PM2.5 in each quantile. Income and housing price data come from CLDS, and PM2.5 data are drawn from the Global Annual PM2.5 Grides.

Table1: Information Disclosure and Income-Pollution Gradient

	Dependent variable: Log Real Income					
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS regression			
PM 2.5 × Disclosure	-0.000245	-0.000397	-0.000363	-0.000236	-0.000411	-0.000440
	(0.00133)	(0.00129)	(0.00134)	(0.00137)	(0.00133)	(0.00136)
PM 2.5	-0.000806	-0.000995	-0.000246	-0.000596	-0.000796	-0.000108
	(0.00326)	(0.00328)	(0.00333)	(0.00316)	(0.00319)	(0.00327)
Disclosure	0.0129	0.0145	0.0229	0.0212	0.0240	0.0323
	(0.0679)	(0.0702)	(0.0721)	(0.0686)	(0.0684)	(0.0711)
Adjusted R-squared	0.416	0.416	0.417	0.416	0.416	0.417
			Panel B: IV regression			
PM 2.5 × Disclosure	0.00836**	0.00838**	0.00870**	0.00862**	0.00856**	0.00868**
	(0.00398)	(0.00408)	(0.00395)	(0.00410)	(0.00419)	(0.00403)
PM 2.5	0.00830	0.00845	0.0141	0.00728	0.00765	0.0127
	(0.0266)	(0.0261)	(0.0297)	(0.0265)	(0.0260)	(0.0295)
Disclosure	-0.341*	-0.352*	-0.349*	-0.342*	-0.347*	-0.340*
	(0.184)	(0.188)	(0.188)	(0.190)	(0.192)	(0.191)
ΔMWTP	167.2	167.5	174	172.3	171.3	173.5
Observations	34,731	34,731	34,731	34,731	34,731	34,731
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. In Panel B, we use instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Standard errors that are clustered at the city level are reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

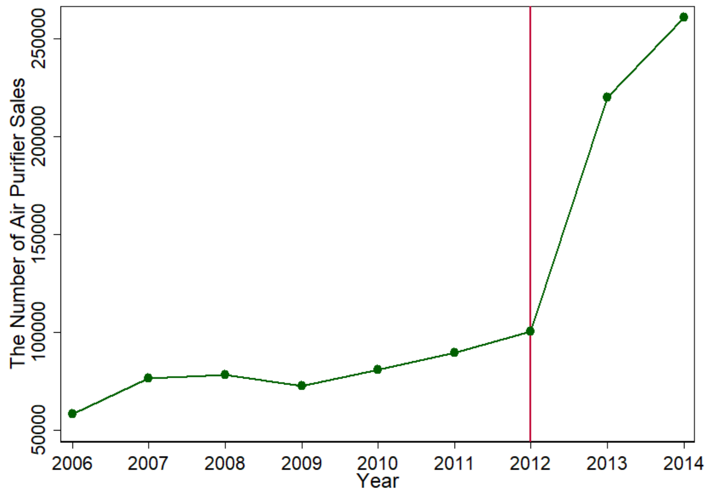
Appendix

Table of Contents

Appendix A: Additional Tables and Figures

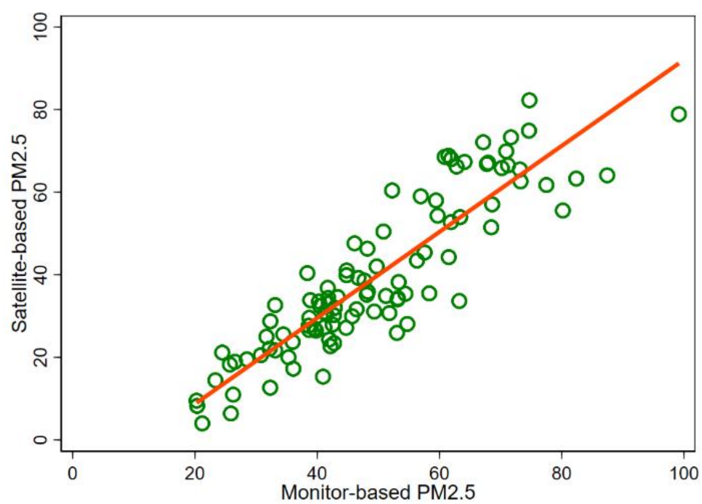
A.1Additional Figures

Figure A1: The number of air purifier sales from 2006 to 2014



Notes: Air purifier transaction data collected by a consulting company.

Figure A2: Monitor-based PM2.5 and Satellite-based PM2.5



Notes: Cities are grouped into one hundred groups according to the quantile of PM2.5 concentration measured by ground monitors. The y-axis denotes the mean value of satellite-based PM2.5 in each quantile, and x-axis denotes the mean value of monitor-based PM2.5 in each quantile. Monitor-based data come from the official website of the MEP of China, and satellite-based PM2.5 data are drawn from the Global Annual PM2.5 Grides.

Figure A3a: Baseline Wind Direction IV

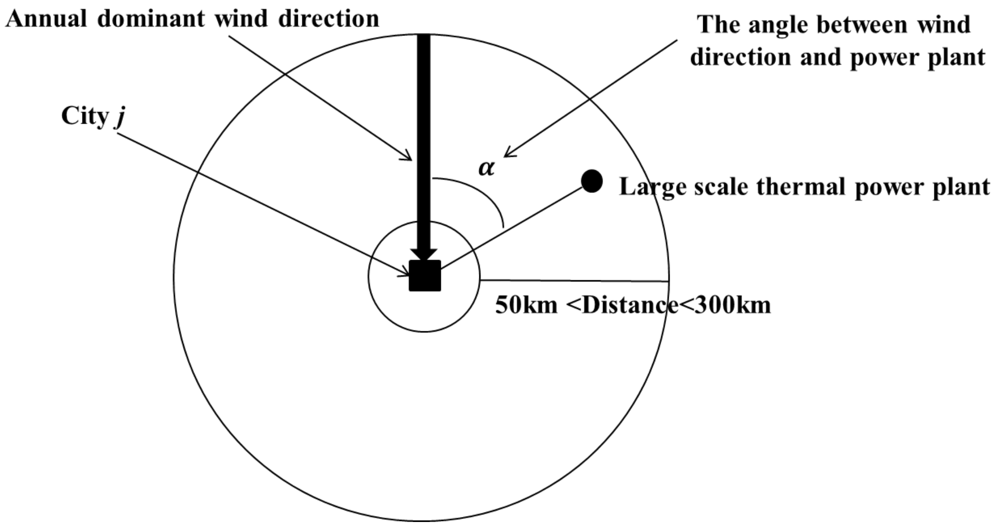
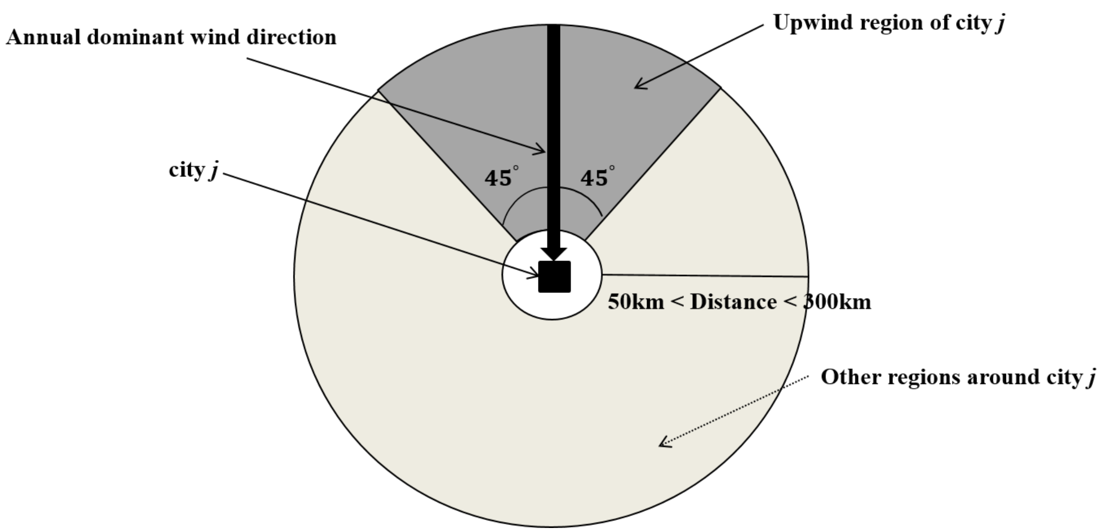
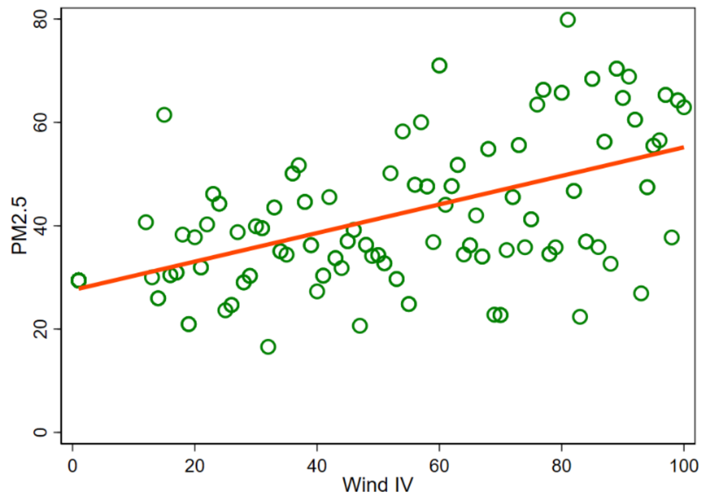


Figure A3b: Alternative Wind Direction IV



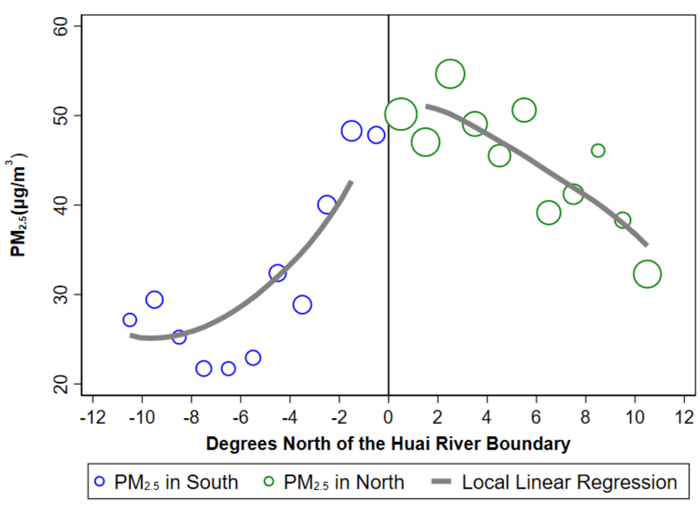
Notes: The thick arrow denotes the annual dominant wind direction of city j. In Figure A2a, the dark dot denotes a large-scale thermal power plant located at least 50km outside city j and within 150km from the city. The angle α represents the angle between the annual prevailing wind direction of city j and the large-scale power plant. Large-scale thermal power plants are defined as plants whose installed-capacities are larger than 1.5 million KW. In Figure A2b, the dark grey area is the upwind area of city j, which is defined as a section of a circular buffer drawn at a distance of 50km-300km from the city, and the angle between the left/right side of the section and the annual dominant wind direction of city j is 45 degree. We exclude the upwind area from the 50km-300km 'loop' around city j, and the light grey area is defined as the non-upwind area around the city.

Figure A4: First Stage Relationship of Baseline Wind direction IV



Notes: Cities are grouped into one hundred groups according to the quantile of the baseline wind direction IV measure. The y-axis denotes the mean value of PM2.5 in each quantile and x-axis denotes the quantile of wind direction IV. PM2.5 data are drawn from the Global Annual PM2.5 Grides.

Figure A5: Discontinuity in PM10 and PM2.5 at the Huai River



Notes: The y-axis denotes PM2.5 concentration, and the x-axis denotes relative latitude north to the river boundary. Cities with positive degrees are located north of the river border, those with negative degrees are located south of the river border. PM2.5 data are drawn from the Global Annual PM2.5 Grides.

A.2 Summary Statistics and Balanced Tests

Table A1: Summary Statistics

Variable name	Description	Mean	Std. dev
Real income	Nominal income adjusted by housing prices	21,459.600	19,880.842
Rural hukou	Indicator = 1 if the person holds rural hukou, =0 otherwise	0.705	0.456
Mid-skill	Indicator = 1 if the highest degree is high school, =0 otherwise	0.178	0.382
High-skill	Indicator = 1 if the highest degree is some college or above, =0 otherwise	0.122	0.327
Male	Indicator = 1 if the person is male	0.483	0.500
Age		44.146	14.192
Pollution Levels	Annual PM2.5 concentration	43.285	17.145
Temperature	Annual mean temperature	156.999	51.409
Humidity	Annual mean humidity	67.936	10.336
GDP per capita		57,152.041	36,090.569
Population	Number of city population	595.341	426.090
Indu elec cons	Industrial electricity consumption	1,154,910.276	1,504,542.415
Total elec cons	Total electricity consumption	1,931,323.405	2,604,252.372
Wastewater	Industrial wastewater emission	9,428.148	9,251.731
SO2	Industrial SO2 emission	55,054.823	59,742.000
Dust	Industrial Dust emission	48,186.896	222,583.024

Note: Table shows summary statistics for outcome variable and most control variables. Real Income is measured by $I_{i,j}/\rho_j^{\theta}$. I and ρ are nominal income and housing price, respectively, and θ is the share of housing expenditure in income. Total electricity consumption includes industrial, residential and commercial consumption.

Table A2: Lagged City Attributes and Data Disclosure

Dependent variable:	Data Disclosure Indicator			
	(1)	(2)	(3)	(4)
Lagged GDP	0.00354	-0.00857	0.00841	-0.0114
	(0.0167)	(0.0126)	(0.0178)	(0.0131)
Lagged Population	0.0149	0.0168	0.0149	0.0173
	(0.0660)	(0.0442)	(0.0655)	(0.0443)
Share of Secondary Industry in GDP	0.00408	0.00897	0.00403	0.00880
	(0.0122)	(0.00889)	(0.0122)	(0.00881)
Share of Tertiary Industry in GDP	0.00165	0.0119	0.00163	0.0117
	(0.0114)	(0.00853)	(0.0113)	(0.00852)
Lagged SO^2 Emission	0.0906	0.0291	0.0877	0.0361
	(0.0634)	(0.0465)	(0.0657)	(0.0432)
Lagged Total electricity consumption	-0.0250	0.0462		
	(0.0987)	(0.0595)		
Lagged Industrial electricity consumption			-0.0314	0.0321
			(0.0546)	(0.0304)
Observations	308	308	308	308
Adjusted R-squared	0.665	0.777	0.779	0.665
City FE	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	Yes	Yes
City-Tier FE × Year FE	No	Yes	No	Yes

Notes: City-level panel regression across 2011, 2013 and 2015. Dependent variable is an indicator whether PM2.5 data have been published in a given city-year, independent variables are city attributes lagged by one year. We use cities that are included in our baseline regression and drop cities with missing values in city attributes. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Changes in Pollution, Wind Direction IV and City Attributes Before and After the Disclosure

Independent variable: Information Disclosure Indicator			
	(1)	(2)	(3)
Panel A. Pollution levels			
PM 2.5	-0.775	-0.616	-0.459
	(0.745)	(0.925)	(0.975)
Panel B. Wind Direction IV			
Baseline Wind Direction IV	-3.427	-5.482	-4.623
	(3.685)	(4.435)	(5.164)
Wind Direction IV excluding Plants within 80km	0.388	-1.929	-0.398
	(3.073)	(3.758)	(4.458)
Wind Direction IV excluding Plants within 130km	-1.034	-3.161	-3.080
	(2.277)	(3.010)	(3.472)
Panel C. Amenities			
Number of Doctors	-501.5	-200.9	95.05
	(522.0)	(346.3)	(386.4)
Area of green coverage	313.8	134.8	297.0
	(351.9)	(508.7)	(547.0)
Panel D. Economic condition			
GDP per Capita	-192.6	-367.5	-673.0
	(529.1)	(421.2)	(433.0)
Share of Secondary Industry in GDP	0.188	-0.386	-0.594
	(0.570)	(0.778)	(0.824)
Share of Tertiary Industry in GDP	0.0233	0.849	1.095
	(0.521)	(0.738)	(0.728)
Panel E. Electricity Consumption			
Industrial Electricity Consumption	-0.294	3.160	3.008
	(4.178)	(5.247)	(5.346)
Total Electricity Consumption	-5.372	4.434	3.970
	(5.963)	(5.618)	(5.772)
Panel F. Emissions			
Industrial SO^2 emission	5.993	3.128	1.500
	(4.646)	(3.637)	(2.898)
Industrial dust emission	130.7	147.3	189.3
	(118.2)	(140.7)	(180.8)
Panel G. Government Expenditure			
Government Public Invest	-25.51	-26.45	-28.71
	(16.24)	(21.37)	(23.53)
Urban Environment Expenditure	-0.485	2.291	-0.274
	(1.568)	(2.160)	(0.536)

Waste Water Treatment Expenditure	-2.329	1.817	-0.302
	(1.683)	(2.955)	(0.639)
City FE	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No

Notes: Row names show the dependent variable. The independent variable is an indicator for whether real time data have been published in a given city. We use cities that are included in our baseline regression and drop cities with missing values in city attributes. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 First-Stage Results

Table A4: First-stage of Wind Direction IV

	Dependent variable: PM 2.5					
	(1)	(2)	(3)	(4)	(5)	(6)
Wind Direction and Coal Plants	0.0315***	0.0321***	0.0298***	0.0323***	0.0329***	0.0303***
	(0.00790)	(0.00820)	(0.00815)	(0.00769)	(0.00794)	(0.00811)
Observations	34,731	34,731	34,731	34,731	34,731	34,731
Adjusted R-squared	0.982	0.982	0.983	0.982	0.982	0.983
F-Statistics	15.89	15.39	13.38	17.58	17.16	17.16
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Demographics include age, gender, hukou status and indicators for education attainment. City controls include temperature and humidity. We use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: First-stage of Huai River RD

	Dependent variable: PM 2.5					
	(1)	(2)	(3)	(4)	(5)	(6)
North × Year 2011	7.362***	7.423***	7.898***	7.230***	7.323***	7.763***
	(1.689)	(1.701)	(1.791)	(1.671)	(1.683)	(1.770)
North × Year 2013	10.45***	10.48***	10.76***	10.36***	10.40***	10.68***
	(2.278)	(2.320)	(2.430)	(2.254)	(2.296)	(2.430)
Control function for the running variable	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE
Observations	20,162	20,162	20,162	20,162	20,162	20,162
Adjusted R-squared	0.986	0.986	0.986	0.986	0.986	0.987
F-Statistics	11.10	10.81	10.69	11.09	10.84	10.63
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE× Year FE	No	No	Yes	No	No	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Longitude Decile× Year FE	No	No	No	Yes	Yes	Yes

Notes: Demographics include age, gender, hukou status and indicators for education attainment. City controls include temperature and humidity. We control for the interaction between the local linear distance to Huai River and year-fixed effects. We use cities that are included in all the three waves of our CLDS sample and located within a 400-mile bandwidth around the river boundary. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Additional Results and Tests of Baseline Wind Direction IV

Table A6: Different Distance Bins for Selection of Power Plants

	Dependent variable: Log real income							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A							
	Baseline Results				Exclude 80km			
PM 2.5 × Disclosure	0.00836**	0.00838**	0.00862**	0.00856**	0.00757*	0.00770*	0.00812*	0.00822*
	(0.00398)	(0.00408)	(0.00410)	(0.00419)	(0.00436)	(0.00449)	(0.00454)	(0.00464)
ΔMWTP	167.2	167.5	172.3	171.3	151.4	154	162.5	164.5
	Panel B							
	Exclude 130km				Exclude 180km			
PM 2.5 × Disclosure	0.00897*	0.00904	0.0101*	0.0102*	0.0118*	0.0117**	0.0123**	0.0123**
	(0.00537)	(0.00551)	(0.00570)	(0.00581)	(0.00600)	(0.00590)	(0.00612)	(0.00603)
ΔMWTP	179.4	180.8	202.4	203.3	235.7	234.7	245.9	245.1
Observations	34,731	34,731	34,731	34,731	34,731	34,731	34,731	34,731
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-Tier FE × Year FE	No	Yes	No	Yes	No	Yes	No	Yes
Weather controls	No	No	Yes	Yes	No	No	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We use individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. We replicate baseline results in column (1)-(4) of Panel A, exclude plants within 80 km around the city in column (5)-(8) of Panel B, exclude plants within 130 km in column (1)-(4) of Panel B, and exclude plants within 180 km in column (5)-(8) of Panel B. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Excluding Newly Built Power Plants

	Dependent variable: Log real income							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A				
	Plants> 5 yrs ago				Plants> 6 yrs ago			
PM 2.5 × Disclosure	0.00764**	0.00760*	0.00786**	0.00774*	0.00788**	0.00795**	0.00819**	0.00819**
	(0.00379)	(0.00386)	(0.00393)	(0.00400)	(0.00385)	(0.00392)	(0.00398)	(0.00405)
ΔMWTP	152.8	152	157.2	154.9	157.7	158.9	163.7	163.8
				Panel B				
	Plants> 7 yrs ago				Plants> 8 yrs ago			
PM 2.5 × Disclosure	0.00839*	0.00850*	0.00901*	0.00907*	0.00775*	0.00778*	0.00830*	0.00830*
	(0.00446)	(0.00452)	(0.00466)	(0.00471)	(0.00417)	(0.00422)	(0.00446)	(0.00452)
ΔMWTP	167.9	170	180.3	181.3	155	155.6	166.1	166.1
Observations	34,731	34,731	34,731	34,731	34,731	34,731	34,731	34,731
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-Tier FE × Year FE	No	Yes	No	Yes	No	Yes	No	Yes
Weather controls	No	No	Yes	Yes	No	No	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Cities affected by new plants included in sample (i.e. in the 'control' regions) so as to generate conservative estimates. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We use individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. We use power plants built more than 5 years ago in column (1)-(4) of Panel A, power plants built more than 6 years ago in column (5)-(8) of Panel A, power plants built more than 7 years ago in column (1)-(4) of Panel B, and power plants built more than 8 years ago in column (5)-(8) of Panel B. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Excluding Coal Producing Region

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
PM 2.5 × Disclosure	0.00864**	0.00873**	0.00888**	0.00887**	0.00890**	0.00887**
	(0.00400)	(0.00410)	(0.00394)	(0.00410)	(0.00418)	(0.00400)
ΔMWTP	172.9	174.6	177.6	177.4	178	177.4
Observations	34,091	34,091	34,091	34,091	34,091	34,091
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We use individual-level pool cross-section data across 2011, 2013 and 2015, use cities that are included in all the three waves of our CLDS sample, and drop all cities in Shanxi Province. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Controlling for Electricity Demand

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Add Industrial Elec. Cons.						
PM 2.5 × Disclosure	0.00773*	0.00789*	0.00864**	0.00777*	0.00789*	0.00848*
	(0.00435)	(0.00444)	(0.00426)	(0.00449)	(0.00458)	(0.00436)
ΔMWTP	154.6	157.7	172.9	155.4	157.7	169.7
Panel B: Add Elec. Cons.						
PM 2.5 × Disclosure	0.00748*	0.00761*	0.00840**	0.00759*	0.00769*	0.00829*
	(0.00431)	(0.00441)	(0.00418)	(0.00446)	(0.00457)	(0.00430)
ΔMWTP	149.6	152.3	168	151.8	153.8	165.7
Observations	32,118	32,118	32,118	32,118	32,118	32,118
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. Total electricity consumption includes industrial, residential and commercial consumption. We use individual-level pool cross-section data across 2011, 2013 and 2015, use cities that are included in all the three waves of our CLDS sample, and drop cities with missing values in electricity consumption. We add Dalh correction terms to account for the potential Roy sorting issue. Total electricity consumption includes industrial, residential and commercial consumption. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10 Baseline Economy and the Wind Direction IV

Dependent variable:	The ratio of upwind plants		Wind direction and coal plants IV	
	(1)	(2)	(3)	(4)
Baseline GDP	-0.00641	-0.00611	0.156	-0.0242
	(0.00889)	(0.00513)	(0.564)	(0.290)
Share of Secondary Industry in GDP	0.00120	0.000793	0.274	0.242
	(0.00394)	(0.00392)	(0.301)	(0.298)
Share of Tertiary Industry in GDP	-0.00288	-0.00275	0.154	0.157
	(0.00590)	(0.00592)	(0.337)	(0.333)
Baseline Population	-0.00744	-0.00734	-0.611	-0.649
	(0.00627)	(0.00601)	(0.480)	(0.468)
Baseline Elec cons	0.000646		0.000145	
	(0.000903)		(0.0615)	
Baseline Industrial Elec cons		0.000949		0.0303
		(0.000773)		(0.0475)
Observations	106	106	106	106
Adjusted R-squared	0.0679	0.0761	0.216	0.218
Region FE	Yes	Yes	Yes	Yes
City-Tier FE	Yes	Yes	Yes	Yes

Notes: City-level regression. Dependent variables are based on power plants built post 2005, and independent variables are measured in the year 2004. We use cities that are included in all the three waves of our CLDS sample, and drop cities with missing values in baseline characteristics. Standard errors clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.5 Results of the Alternative Wind Direction IV

Table A11: First-stage of Alternative Wind Direction IV

	Dependent variable: PM 2.5					
	(1)	(2)	(3)	(4)	(5)	(6)
Upwind coal consumption	0.0193***	0.0193***	0.0173***	0.0198***	0.0198***	0.0175***
	(0.00582)	(0.00596)	(0.00577)	(0.00564)	(0.00577)	(0.00564)
Observations	34,731	34,731	34,731	34,731	34,731	34,731
Adjusted R-squared	0.982	0.982	0.983	0.982	0.982	0.983
F-value	10.99	10.51	8.993	12.26	11.79	11.79
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE× Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Upwind coal consumption denotes the total coal consumption of power plants located at the upwind area of a given city. We control for the total coal consumption of plants at the counterpart non-upwind area of the city. We use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We use cities that are included in all the three waves of our CLDS sample. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Alternative Wind Direction IV with Distance Bins for Selection of Plants

	Dependent variable: Log real income							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A				
	Baseline Results				Exclude 80km			
PM 2.5 × Disclosure	0.0112**	0.0112**	0.0112**	0.0109**	0.0122**	0.0122**	0.0125**	0.0124**
	(0.00486)	(0.00521)	(0.00519)	(0.00548)	(0.00568)	(0.00584)	(0.00605)	(0.00619)
ΔMWTP	223.7	223.1	223.2	218.2	243.4	243.1	250.8	248.9
				Panel B				
	Exclude 130km				Exclude 180km			
PM 2.5 × Disclosure	0.0111*	0.0110*	0.0121*	0.0119	0.00868**	0.00862**	0.00875*	0.00869*
	(0.00617)	(0.00646)	(0.00700)	(0.00727)	(0.00427)	(0.00431)	(0.00459)	(0.00464)
ΔMWTP	222.7	219.5	241.9	237.4	173.7	172.3	175	173.9
Observations	34,731	34,731	34,731	34,731	34,731	34,731	34,731	34,731
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-Tier FE × Year FE	No	Yes	No	Yes	No	Yes	No	Yes
Weather controls	No	No	Yes	Yes	No	No	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the total coal consumption of power plants located at the upwind area of a given city, controlling for the total coal consumption of plants at the counterpart non-upwind area of the city. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. We exclude plants within 50 km around the city in column (1)-(4) of Panel A, exclude plants within 80 km in column (5)-(8) of Panel B, exclude plants within 130 km in column (1)-(4) of Panel B, and exclude plants within 180 km in Column (5)-(8) of Panel B. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: The Coal-fired Plants Located Upwind of Large Metropolitans

City	Number of Upwind Plants	Ratio of Upwind Plants	Coal Consumption of Upwind Plants	Smallest Angle of Plants
Beijing	3	37.5%	194.779	22.816
Tianjin	3	37.5%	149.590	15.368
Shanghai	3	13.6%	105.619	3.347
Guangzhou	5	62.5%	170.603	0.088
Shenzhen	4	57.1%	152.005	1.184
National mean	2	35.8%	79.161	17.194

Notes: The statistics are calculated using the large-scale thermal power plants located at 50-300km from a given city. We define the upwind area as a section of a circular buffer drawn at a distance of 50-300km from a given city, and the angle between the section and the annual dominant wind direction of the city being at least 45 degree.

A.6 Results of Additional Controls and Alternative Samples

Table A14: Additional Controls: Economic Condition, Amenities, and Emissions

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Add Economic Controls			
PM 2.5 × Disclosure	0.00825*	0.00895*	0.00902**	0.00817*	0.00883*	0.00881**
	(0.00448)	(0.00475)	(0.00417)	(0.00470)	(0.00492)	(0.00431)
ΔMWTP	165.1	179	180.4	163.5	176.5	176.1
Observations	32,766	32,766	32,766	32,766	32,766	32,766
			Panel B: Add Other Amenities			
PM 2.5 × Disclosure	0.00828*	0.00850*	0.00861*	0.00811*	0.00832*	0.00846*
	(0.00447)	(0.00453)	(0.00436)	(0.00461)	(0.00465)	(0.00446)
ΔMWTP	165.6	170	172.2	162.2	166.3	169.2
Observations	32,400	32,400	32,400	32,400	32,400	32,400
			Panel C: Add Emission Controls			
PM 2.5 × Disclosure	0.00765*	0.00792*	0.00839*	0.00778*	0.00803*	0.00838*
	(0.00442)	(0.00451)	(0.00442)	(0.00456)	(0.00463)	(0.00450)
ΔMWTP	153	158.4	167.7	155.6	160.6	167.7
Observations	31,263	31,263	31,263	31,263	31,263	31,263
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. Economic Controls include GDP per capita, the share secondary industry in GDP and share of tertiary industry in GDP. Other Amenities include the number of doctors, the number of library books and the area of green coverage. Emission Controls include industrial water emission, industrial SO2 emission and industrial dust emission. We employ individual-level pool cross-section data across 2011, 2013 and 2015, use cities that are included in all the three waves of our CLDS sample, and drop cities with missing values in these control variables. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Excluding Big Cities and High Polluters

	Dependent variable: Log real income					
City Excluded	Beijing	Tianjin	Shanghai	Shenyang	Zhengzhou	Wuhan
	(1)	(2)	(3)	(4)	(5)	(6)
PM 2.5 × Disclosure	0.00836**	0.00862**	0.00858**	0.00841**	0.00828**	0.00860**
	(0.00418)	(0.00428)	(0.00422)	(0.00418)	(0.00414)	(0.00402)
ΔMWTP	167.1	172.4	171.6	168.1	165.7	171.9
Observations	34,314	34,341	34,368	34,568	34,380	34,196
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
City-Tier FE × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.7 Results of Huai River Regression Discontinuity

Table A16: Huai River Regression Discontinuity

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
PM 2.5 × Disclosure	0.0158*	0.0164*	0.0157*	0.0157*	0.0163*	0.0158*
	(0.00816)	(0.00868)	(0.00825)	(0.00837)	(0.00867)	(0.00830)
PM 2.5	0.0177	0.0167	0.0114	0.0195	0.0183	0.0125
	(0.0155)	(0.0152)	(0.0137)	(0.0153)	(0.0151)	(0.0133)
Disclosure	-0.850*	-0.916*	-0.919*	-0.860*	-0.927*	-0.960**
	(0.460)	(0.495)	(0.472)	(0.477)	(0.499)	(0.476)
ΔMWTP	316.7	327.8	313.4	314.4	325.1	316.8
Control function for the running variable	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE
Observations	20,162	20,162	20,162	20,162	20,162	20,162
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Longitude Quintile FE × Year FE	No	No	No	Yes	Yes	Yes

Notes: Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We control for the interaction between the local linear distance to Huai River and year-fixed effects. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample and located within a 400-mile bandwidth around the river boundary. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Alternative Function for the Running Variable

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
PM 2.5 × Disclosure	0.0159*	0.0165*	0.0156*	0.0158*	0.0164*	0.0158*
	(0.00847)	(0.00901)	(0.00842)	(0.00870)	(0.00901)	(0.00848)
PM 2.5	0.0213	0.0202	0.0139	0.0232	0.0218	0.0150
	(0.0164)	(0.0162)	(0.0141)	(0.0163)	(0.0160)	(0.0137)
Disclosure	-0.858*	-0.929*	-0.921*	-0.865*	-0.937*	-0.963*
	(0.475)	(0.512)	(0.478)	(0.493)	(0.516)	(0.483)
ΔMWTP	319	331	311.9	315.5	327.2	315
Control function for the running variable	Quadratic × Year FE	Quadratic × Year FE	Quadratic × Year FE	Quadratic × Year FE	Quadratic × Year FE	Quadratic × Year FE
Observations	20,162	20,162	20,162	20,162	20,162	20,162
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Longitude Quintile FE × Year FE	No	No	No	Yes	Yes	Yes

Notes: Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We control for interactions between quadratic controls for the distance to Huai River and year-fixed effects. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample and located within a 400-mile bandwidth around the river boundary. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TableA18 Various Bandwidths in Regression Discontinuity

	Dependent variable: Log real income			
	(1)	(2)	(3)	(4)
	Panel A: 500 Miles			
PM 2.5 × Disclosure	0.0173**	0.0183*	0.0185**	0.0197**
	(0.00865)	(0.00933)	(0.00905)	(0.00977)
Observations	23,050	23,050	23,050	23,050
	Panel B: 525 Miles			
PM 2.5 × Disclosure	0.0239*	0.0253*	0.0248*	0.0263*
	(0.0121)	(0.0135)	(0.0125)	(0.0139)
Observations	24,030	24,030	24,030	24,030
	Panel C: 550 Miles			
PM 2.5 × Disclosure	0.0165*	0.0173*	0.0178*	0.0189*
	(0.00924)	(0.00992)	(0.00994)	(0.0109)
Observations	24,935	24,935	24,935	24,935
Control function for the running variable	Linear × North× Year FE	Linear × North× Year FE	Quadratic× Year FE	Quadratic× Year FE
City FE	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	Yes	Yes
City-Tier FE × Year FE	No	Yes	No	Yes
Weather controls	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes

Notes: Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. In column (1), (3) and (5), we control for the interaction between the local linear distance to Huai River and year-fixed effects. In column (2), (4), and (6), we control for the interactions between quadratic controls for the distance to Huai River and year-fixed effects. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample and located within various bandwidths around the river boundary. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.8 Results of Placebo Tests

Table A19: Placebo Wind Directions

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
PM 2.5 × Disclosure	0.00107	0.00109	0.00162	0.00290	0.00296	0.00237
	(0.00700)	(0.00737)	(0.00729)	(0.0107)	(0.0110)	(0.00919)
Observations	34,731	34,731	34,731	34,731	34,731	34,731
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. We use 'placebo' wind direction by adding 180 degree. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. We add 180 degree to wind direction angle. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A20: Placebo RD Cutoff

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
PM 2.5 × Disclosure	-0.0308	-0.0296	-0.0261	-0.0326	-0.0310	-0.0270
	(0.0370)	(0.0347)	(0.0236)	(0.0391)	(0.0361)	(0.0237)
PM 2.5	0.00620	0.00594	-0.00175	0.00738	0.00697	-0.000820
	(0.0234)	(0.0227)	(0.0200)	(0.0242)	(0.0232)	(0.0198)
Disclosure	1.609	1.543	1.425	1.705	1.615	1.468
	(1.879)	(1.814)	(1.290)	(1.986)	(1.888)	(1.300)
Control function for the running variable	Linear × North× Year FE	Linear × North× Year FE	Linear × North× Year FE	Quadratic× Year FE	Quadratic× Year FE	Quadratic× Year FE
Observations	18,136	18,136	18,136	18,136	18,136	18,136
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We move the river border parallelly by 5 degree. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample and located within a 400 mile-bandwidth around the 'placebo' river boundary. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A21: Placebo Data Rollout Sequence

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
PM 2.5 × Disclosure	0.00764	0.00745	0.00764	0.00908	0.00886	0.00892
	(0.00553)	(0.00576)	(0.00679)	(0.00576)	(0.00595)	(0.00722)
PM 2.5	-0.0153	-0.0142	-0.0109	-0.0186	-0.0174	-0.0148
	(0.0346)	(0.0350)	(0.0430)	(0.0348)	(0.0353)	(0.0445)
Disclosure	-0.394	-0.384	-0.405	-0.456*	-0.445*	-0.458
	(0.249)	(0.259)	(0.300)	(0.258)	(0.267)	(0.318)
Observations	34,731	34,731	34,731	34,731	34,731	34,731
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We randomly allocate cities to each of the three waves of data disclosure. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A22: Placebo Data Rollout Time

	Dependent variable: Log real income					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Delay One Year					
PM 2.5 × Disclosure	0.00972	0.00968	0.00654	0.0110	0.0110	0.00784
	(0.00753)	(0.00768)	(0.00889)	(0.00796)	(0.00814)	(0.00915)
	Panel B: Delay Two Years					
PM 2.5 × Disclosure	0.00377	0.00357	0.00278	0.00511	0.00490	0.00394
	(0.00345)	(0.00349)	(0.00437)	(0.00352)	(0.00350)	(0.00452)
Observations	34,731	34,731	34,731	34,731	34,731	34,731
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE × Year FE	Yes	Yes	No	Yes	Yes	No
City-Tier FE × Year FE	No	Yes	No	No	Yes	No
Region FE × City-Tier FE × Year FE	No	No	Yes	No	No	Yes
Weather controls	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instrumental variables specification using the interaction between wind direction, the location and the coal consumption of power plants. Demographics include age, gender, hukou status and indicators for education attainment. Weather controls include temperature and humidity. We employ individual-level pool cross-section data across 2011, 2013 and 2015, and use cities that are included in all the three waves of our CLDS sample. We add Dalh correction terms to account for the potential Roy sorting issue. Standard errors that are clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B: City-Specific Housing Price Measure

We define the value of the home occupied by individual i in city j, $P_{i,j}$, as the value of the house (for owner-occupied units) or annual rent (for rental units) in our census data. Following Bayer et al. (2007, 2009), we assume that $P_{i,j}$ is a function of a scaling parameter ρ_j specific to city j and a vector of housing characteristics H_i. Then, we estimate the following regression:

$$Log\ P_{i,j,t} = Log\rho_{j,t} + \phi_t \Omega_{i,owner} + H_i'\tau_t + \varepsilon_{i,j,t}^H \tag{B1}$$

Where $\Omega_{i,owner}$ is a dummy variable that takes the value of 1 for a owner-occupied unit and 0 otherwise. Dwelling characteristics H_i describe the number of rooms, floor area, whether tap water is provided, whether a kitchen is provided, whether a shower room is provided and whether there is a bathroom. We run the housing price regression of Equation (B1) for 2011, 2013 and 2015, separately. The city-and year-specific fixed effect $Log\rho_{jt}$ measures the effective "price of housing services" in city j and year t independent of ownership and housing attributes. Therefore, our housing price measure captures the variation in housing costs both spatial and time variations in housing costs.

Appendix C: Roy Sorting and the Dahl Correction Approach

Roy sorting refers to the problem that individuals respond to idiosyncratic wage draws and are likely to move to a location where that wage draw is good. For instance, individuals from a particular region could earn unusual high wages in a given place, because their personal abilities have unusual comparative advantages specific for working in this place. Thus, other people who look like these individuals cannot earn same wages if they move to the place. We follow the semi-parametric approach proposed by Dahl (2002) to address the Roy sorting bias.

We use population census data in 2005 and divide individuals into groups based on their original regions. Within each original region cell, we also allocate individuals into one of the two education classes: high school dropouts and high school graduates. Then, we define selection probability ω_i as the fraction of the population in individual $i's$ cell that chooses to live in a particular destination region. Finally, we augment wage hedonic equation by adding a quadratic function of ω_i. Controlling for selection probability can effectively correct for Roy sorting bias (Dahl, 2002).

We use baseline year data in 2005 to measure the selection probability, and control for city-fixed effects in our baseline empirical specification of equation (10). Therefore, we can account for any location-specific unobservables that may be correlated with these Dahl correction terms.

The selection probability ω_i can also capture bilateral migration costs across regions. If there are high mobility costs associated with moving form region A to B, the share of people moving from region A to B should be small (Bryan and Morten 2019).