Tsukuba Economics Working Papers
No. 2024-001

Does Exposure to PM2.5 Increase the Likelihood of Early Retirement in Middle-Aged Individuals? Evidence from Chinese Data

by

Meiyi Zhuang

Xinyi Zhang

Hisahiro Naito

May 2024

UNIVERSITY OF TSUKUBA
Faculty of Humanities and Social Sciences
1-1-1 Tennodai
Tsukuba, Ibaraki 305-8571
JAPAN

Does Exposure to PM2.5 Increase the Likelihood of Early Retirement in Middle-Aged Individuals? Evidence from Chinese Data

Meiyi Zhuang *†
Program in Economic and Public Policy
Graduate School of Humanities and Social Sciences
University of Tsukuba

Xinyi Zhang *‡
Graduate School of Humanities and Social Sciences
University of Tsukuba

Hisahiro Naito *§
Graduate School of Humanities and Social Sciences
University of Tsukuba

May 2024

*The authors appreciate the comments from Taiyo Fukai, Kennnichi Kashiwagi, Yoshinori Kurokawa, Mohammad Abdul Malek, Midori Matsuhima,Yuko Nakano and Zheng Fei Yu. The authors are responsible for all remaining errors. H.N. appreciate financial support from the university research fund and from Japan Society of Promotion of Sciences KAKENHI Grant 23H00827.

†Email: zmeiyi95@gmail.com

‡Email: mint.weixin12@gmail.com

§Corresponding author. Email: naito@dpipe.tsukuba.ac.jp. Address: Tennodai 1-1-1, Tsukuba City, Ibaraki Prefecture, Japan, Postal Code 305-0032

Abstract

Health is one of the most critical factors that affects retirement behavior, and poor health may lead to early retirement among middle-aged and older adults. In China, where the population is aging rapidly, early retirement has significant implications for the economy. Recent studies have shown that air pollution, particularly PM2.5, can cause various illnesses, such as respiratory diseases, cardiovascular diseases, high blood pressure, and diabetes. In this paper, we analyze the effects of PM2.5 on the retirement and health of middle-aged and elderly people, assuming that the effects of air pollution on retirement are highly nonlinear and different for farmers and non-farmers. To control for potential endogeneity, we use 2SLS estimation.

The regression results for non-farmers show that higher PM2.5 concentrations increase the probability of heart-related diseases and early retirement behavior. Specifically, we found that a 10 microgram/m^3(about one standard deviation) per cubic meter increase in PM2.5 concentration is associated with a 58% increase in the probability of heart-related diseases and a 57% increase in early retirement. This implies that roughly 12.1 million people could continue participating in the labor market if the government can reduce PM2.5 concentration by 10 microgram/m^3per cubic meter across the country. For farmers, we found that higher PM2.5 concentration is associated with a higher probability of lung-related diseases, but we did not find evidence that it increases early retirement. For both non-farmers and farmers, we did not find evidence that a higher PM2.5 concentration decreases financial wealth.

These findings suggest that higher air pollution deteriorates the health of non-farmers, increases the disutility of work, and induces early retirement but does not affect the financial wealth of farmers and non-farmers.

1 Introduction

Many developed countries and some middle-income countries are facing problems related to aging, such as increasing social security expenditures. However, the negative effects of aging can potentially be mitigated if older individuals continue to work. Therefore, retirement decisions among the elderly have significant policy implications for the economies of those countries.

The decision to retire is complex and influenced by various factors, including income, family, work opportunities, and leisure activities. Several empirical studies have highlighted the importance of health as a factor in retirement decisions, although there is no consensus on the size of its effect(Bound, Schoenbaum, Stinebrickner and Waidmann, 1999; Dwyer and Mitchell, 1999; McGarry, 2004; Disney, Emmerson and Wakefield, 2006; Au, Crossley and Schellhorn, 2005; Jones, Rice and Roberts, 2010; Bound, Stinebrickner and Waidmann, 2010; Blundell, Britton, Dias and French, 2023). If an individual's health is poor, they might not be able to continue working as they approach middle age. Furthermore, due to the decline of specific bodily functions, older and middle-aged individuals may be more vulnerable to external factors than younger individuals (Cakmak et al., 2007; Xue et al., 2018).

In the literature on health economics and environmental sciences, air pollution has received significant attention as an important factor affecting people's health and mortality. Numerous studies have shown that high levels of PM2.5 concentration, a common air pollutant, can cause various illnesses, including respiratory diseases, cardiovascular diseases, high blood pressure, and diabetes (Sun, Yue, Ying, Cardounel, Brook, Devlin, Hwang, Zweier, Chen and Rajagopalan, 2008; Andersen, Raaschou-Nielsen, Ketzel, Jensen, Hvidberg, Loft, Tjønneland, Overvad and Sørensen, 2012; Liang, Zhang, Zhao, Ruan, Lian and Fan, 2014; Franklin, Brook and Pope III, 2015; Lin, Guo, Zheng, Di, Liu, Xiao, Li, Zeng, Cummings-Vaughn, Howard et al., 2017). This is because the size of PM2.5 is very small and they do not stop at bronchus when they are inhaled. They are absorbed in bloodstream and cause inflammation in various organs of human body.

According to statistics from the World Health Organization (WHO) on PM2.5 concentration between 2000 and 2013, the annual average value of PM2.5 concentration in most urban areas in China exceeds 40 µg/m^3. This is far beyond the WHO guideline value of 10.0 µg/m^3 (World Health Organization and European Centre for Environment, 2021). In 2015, seven of the world's top ten most polluted cities were located in China (Zhou et al., 2015). The severe situation of air pollution in China unavoidably affects older people since they spend more time outside in their lifetime. Furthermore, middle-aged and older adults are more susceptible to outdoor air pollution compared to younger people (Cakmak et al., 2007; Xue et al., 2018).

On the other hand, China, which has the largest population in the world, is facing the challenge of an aging population (Fang et al., 2015). Over the past two decades (2001-2021), China's population has grown at an average annual rate of 0.53%. During this time, the proportion of adults aged 60 and older has increased by 5.44%, from 13.26% in 2010 to 18.70% in 2020 (National Bureau of Statistics of China, 2011, 2021).

Due to the high level of PM2.5 and aging population in China and the potential impact of health on retirement decisions, a relevant policy question is to what extent air pollution affects people's decision to retire. This research question holds significance not only for China but also for other developing countries, particularly in Asia and Central Africa, where air pollution is a significant problem.

From a theoretical standpoint, the effect of air pollution on retirement is not straightforward. First, people's health might be negatively affected by poor air quality, which can decrease productivity even when individuals are young (He, Liu and Salvo, 2019). If this is the case, it can reduce the lifetime wealth of such individuals. When people's lifetime wealth becomes lower, they might not have sufficient financial resources to retire early. In such a scenario, air pollution might not increase early retirement (income effect). Second, air pollution might not affect the human capital of young individuals but could increase the disutility of work at an older age. Evidence shows that air pollution increases the probability of having chronic diseases such as hypertension and diabetes (Liang et al., 2014; Du et al., 2016). Thus, older people may experience significant physical and psychological difficulties in working if their health is negatively affected (disutility effect). Therefore, the income effect and disutility effect work in opposite directions. This suggests that it is important to examine empirically whether and to what extent air pollution affects the retirement and health behavior of older people.

In this paper, we examine the effects of PM2.5 on the retirement and health of middle-aged and elderly individuals. To achieve this, we combine satellite MERRA-2 data of PM2.5 and micro-data from CHARLS, utilizing city information from CHARLS and residential information of respondents prior to the survey years.

In the literature, there are several studies that examine the effect of air pollution on labor supply (Ostro, 1983; Hausman et al., 1984; Pönkä, 1990; Hanna and Oliva, 2015; Aragon et al., 2017; Huang et al., 2021; Li and Li, 2022; Wu et al., 2023). However, our study is different from the previous studies from several aspects. First, our study focus on the effect of air pollution on retirement while previous studies focus on the labor supply. The retirement decision is a decision to exit from labor market while the labor supply decision is a contemporaneous decision. Second, the effect of air pollution on retirement decision is highly nonlinear function of age while the labor supply decision is not likely. In our specification, we allow the marginal effect of air pollution is a quadratic function of age. Third, the elderly people are more vulnerable to air pollution than young people and the disutility of labor can be higher for old people than for young people. Thus, it is not possible to extrapolate the estimated obtained by using the sample of young people to the sample of old people. To the best of our knowledge, our study is the first paper to examine the effect of air pollution on the retirement decision.

One concern of our estimates is that our estimates are biased due to hidden industry characteristics even when we use 2SLS estimation. We argue that our estimates are likely to biased downward because our estimation results in Table 4-Table 9 show that as we add more city characteristics, the estimated coefficient becomes bigger. If our estimates are mainly driven by hidden city characteristics, the estimated coefficients should be smaller as we add more city characteristics as control variables. This indicates that the bias due to hidden city characteristics is likely to be very small.

The paper is organized as follows: Section 2 summarizes the related literature. Section 3 covers the model and the dataset. Sections 4.1 and 4.2 present the summary statistics and estimation results using the ordinary least squares (OLS) estimation method. In Section 4.3, we present the regression results using the two-stage least squares (2SLS) estimation method. Additionally, Section 4.4 discusses the potential mechanism, and Section 4.5 addresses the potential estimation bias due to migration. Section 5 provides a discussion of the results and draws conclusions.

2 Literature Review

2.1 Health and Retirement

Various studies have examined the importance of health on retirement decision. The results of numerous earlier investigations have varied in this area. Some study emphasize the importance of health on retirement decision while other studies argue that health is not an important factor.

According to a study by McGarry (2004), bad health has a considerably more significant impact on the likelihood of continuing to work than financial factors, even when there are no incentives. Dwyer and Mitchell (1999) found that health issues have a larger impact on retirement planning than economic factors. Through a survey on older adults' contentment with retirement, Bender (2012) discovered that economic factors, health, and voluntary retirement are essential drivers of older adults' retirement satisfaction. Among these, economic factors are of utmost importance.

Disney et al. (2006) examined how lousy health affects people's decisions to retire in the UK using the first eight waves of the British Household Panel Survey. Adverse personal health shocks powerfully predict individual retirement behavior. Clarke et al. (2012) show that self-reported chronic health issues are linked to a reduced likelihood of continuing to work full-time after the age of 62. Men with high career aspirations but not working full-time beyond age 62 were less content with their lives despite controlling for various factors, including health. On the other hand, no such link was identified among women. De Wind et al. (2013) point out persons in good health seek to maximize their quality of life. As a result, they may choose to retire early. This also explains why several earlier quantitative research that looked at the association between health and early retirement failed to find a significant one. Using cross-sectional data from Finland, Ilmakunnas and Ilmakunnas (2018) show that the actual retirement age is less responsive to health.

2.2 Air pollution and Health

Numerous studies in the past have shown that air pollution affects people's health. Ying et al. (2014) showed that long-term exposure to PM2.5 raises blood pressure by stimulating the sympathetic nervous system, which may result in hypothalamus inflammation, through research on rats. Bell et al. (2014) show that there is a statistically significant link between hospitalization for respiratory and cardiovascular conditions and PM2.5 levels. In addition, several other prevalent chronic diseases like diabetes and hypertension are linked to air pollution. Krämer et al. (2010) find that air pollution from traffic is linked to a higher prevalence of type 2 diabetes in older women.

Several studies examine the effect of air pollution for elderly people. Kan et al. (2008) show that air pollution is a crucial aspect affecting older individuals' health. Cakmak et al. (2007) suggest that the very elderly are particularly susceptible to dying from air pollution. Xue et al. (2018) show that respiratory illness in older individuals in Shenyang was substantially correlated with PM10 and NO2 levels.

2.3 Air Pollution and Labor Supply

As we just have discussed in the introduction, there are several studies that examine the effect of air pollution on labor supply (Ostro, 1983; Hausman et al., 1984; Pönkä, 1990; Hanna and Oliva, 2015; Aragon et al., 2017; Huang et al., 2021; Gu et al., 2020; Li and Li, 2022; Wu et al., 2023). For example, Hausman et al. (1984) find that, controlling for city-specific effects, one standard deviation increase in particulates in the two weeks prior is associated with about a 10% increase in lost work days.

Hanna and Oliva (2015) examine the relationship between air pollution and labor supply. They utilized a natural experiment that arose from the closure of a major refinery in Mexico City to explore the causal relationship between pollution and labor supply. The refinery's closure resulted in a 19.7% reduction in pollution, as measured by SO2, in the surrounding neighborhoods. They found that the closure led to an increase in work hours per week of 1.3 hours, representing a 3.5% increase.

Huang, Zhang, Chen and Ning (2021) analyze the effect of PM2.5 on the current labor supply using the China Health and Retirement Longitudinal Survey (CHARLS). They pooled four waves of data and examined the labor supply with the sample selection equation. [1] Note that

[1] To estimate the sample selection equation, it is necessary to have an exclusion variable that affects labor market participation but not labor supply. In this study, the researchers used a dummy variable indicating whether the respondent's age was over 60, assuming that this variable only affected labor market participation. However, since aging can impact both labor market participation and labor supply, this assumption may not hold.

retirement decision is qualitatively different from labor supply in several dimensions. Retirement is the exist from the labor market while the labor supply is contemporaneous decision. Li and Li (2022) examine the effect of perceived air pollution on labor supply in China. Gu, Bian and Elahi (2020) study the effect of air pollution on female labor supply in China.

3 Model and Data Set

3.1 Empirical Model

For modeling the retirement behavior, We believe that the marginal effect of PM2.5 is different for the different age and it is a nonlinear function of age. Thus, we consider the following model:

$$\begin{aligned} Y_{pci} &= \beta_0 + \beta_1(PM2.5_{pc} - \overline{PM2.5}) + \beta_2(Age_{pci} - 45) \times (PM2.5_{pc} - \overline{PM2.5}) \\ &+ \beta_3(Age_{pci} - 45)^2 \times (PM2.5_{pc} - \overline{PM2.5}) + \beta_4(Age_{pci} - 45) + \beta_5(Age_{pci} - 45)^2 \\ &+ \alpha X1_{pci} + \gamma X2_{pc} + \delta X3_p + \varepsilon_{pci} \end{aligned} \tag{1}$$

where p is an index of provinces, c is an index of cities and i is an index of individuals. The variable Y_{pci} represents the outcome, which could be a dummy variable indicating retirement or chronic diseases related to PM2.5 exposure of an individual i in city c. The independent variable $PM2.5_{pc}$ denotes the average PM2.5 concentration in the city where individual i resided between 2004 and 2013. The national average of PM2.5 in China during the same period is denoted as $\overline{PM2.5}$. The age of individual i is denoted as Age_{pci}, while $X1_{pci}$ represents individual characteristics such as gender, marital status, hukou, and year of schooling. In China, hukou is a legal document used to record and retain basic information about the household population, divided into agricultural and non-agricultural hukou. $X2_{pc}$ represents city characteristics, including GDP per capita, the share of secondary and tertiary industries in GDP, average wage, total population, population density, and urban area share. $X3_p$ represents province characteristics, including GDP per capita, the share of secondary and tertiary industries in GDP, total population, urban population density, and average education years.

Additionally, if air pollution is correlated with economic activity and labor demand, it could create an endogeneity problem.

It is noteworthy that the conditional expectation of the derivative of Y_{ic} with respect to $PM2.5_c - \overline{PM2.5}$ is expressed as:

$$E\left[\frac{\partial Y_{pci}}{\partial(PM2.5_{pc} - \overline{PM2.5})}\right] = \beta_1 + \beta_2(Age_{pci} - 45) + \beta_3(Age_{pci} - 45)^2 \tag{2}$$

Thus, β_1 represents the marginal effect of PM2.5 at age 45. The coefficients β_2 and β_3 determine to what extent such a marginal effect of PM2.5 becomes a nonlinear function of age.

3.2 Data Set

We use China Health and Retirement Longitudinal Survey (CHARLS). We downloaded the data and questionnaire from CHARLS' official website on request. The World Bank, the National Institute on Aging's Behavioral and Social Research Division, the National Natural Science Foundation of China, and Peking University all support CHARLS. A longitudinal study of adults in China over 45 is called CHARLS. It seeks to comprehend the socioeconomic causes and effects of aging. The study asks a wide range of questions about older adults' social networks, economic status, and physical and mental health.

From May 2011 to March 2012, the CHARLS Baseline Survey (2011 wave) was conducted across the country in 28 provinces and municipalities. In 150 counties and districts, the survey was conducted in 450 PSUs (village or community). Of the 450 PSUs, 52.67 % were rural and 47.33 % urban. The sample population was chosen as part of a stratified, multi-stage probabilistic design (Zhao et al., 2013). All county-level units, except Tibet, were sampled using proportional probability sampling (PPS), which was then stratified by region, county characteristics (urban vs. rural), and gross domestic product (GDP) per capita. One of the residents who were older than 40 and lived in the sampling household was chosen at random. The selected individual became the respondent if they were 45 or older. 17,708 respondents in 10,257 households in 450 villages and urban communities, 150 districts, and 28 provinces made up this initial sample. In addition, they include the sample whose age is 43 or older to increase the sample size. From July 2013 to January 2014, the second survey wave (2013 wave) was carried out. Up till now, data from four survey waves have been made available. 92 % of rural and 83 % of urban households participating in the first wave completed at least one module in 2013 wave. 87 % and 86 % of the households finished at least one module in waves 3 and 4, respectively. Compared to other household surveys, the success rate is high (Zhao et al., 2020).

To conduct a cross-sectional analysis, we use data from 2011 wave and 2013 wave. We use the respondents who answered both two waves. We choose the respondents whose age are between 43 and 78 in the 2011 wave. Additionally, those who have never worked before 2011 were excluded from the data set.

Individuals often choose a retirement age based on their occupation. Unlike non-farmers, farmers typically do not have a defined retirement age. To distinguish between farmers and non-farmers, we classify individuals as farmers if they worked in agriculture in the previous year or if their last job before retirement was in agriculture. Conversely, if an individual only held non-farm jobs in the recent year or their last employment before retirement was outside of agriculture, they are considered non-farmers.

Retirement refers to a person's complete withdrawal from the labor market. If a person's annual working time is less than 200 hours, we classify them as partially retired from the labor market. If an individual is in partial retirement from the labor market in two consecutive waves, we define them as retired.

For individual characteristics such as education, age, and marital status, and chronic diseases, we use information from the 2013 wave.

Regarding PM2.5-related diseases, we define hypertension, diabetes, chronic lung diseases, heart attack, and asthma as the chronic diseases related to PM2.5 (Andersen et al., 2011; Choi & Cho, 2013; F. Huang et al., 2017; Yeatts et al., 2007). Since these are chronic diseases, we use the second round survey. For example, if a respondent states that they have a heart-related disease in the 2nd round survey, we assume that this person developed a heart-related disease and assign a value of one to the heart-related disease dummy variable.

Regarding PM2.5 concentration information, following the literature on the effect of PM2.5 on health and other outcomes, MERRA-2 in the Goddard Earth Observing System Model (GEOS) version 5.12.4 was used to generate PM2.5 data (Deschenes et al., 2020; Chen et al., 2022). The use of Merra-2 is attractive from its longest span of the years covered. MERRA-2 is available at a 50 $\times$ 60-km grid level for each month since 1980. MERRA-2 is the most recent iteration of the Global Atmospheric Reanalysis for the Satellite Era published by NASA's Global Modeling and Assimilation Office (GMAO). The collection includes assimilated aerosol diagnostics such as complete extinction (and scattering) aerosol optical thickness (AOT) at 550 nm, the surface mass concentration of aerosol components, and column mass density of aerosol components (black carbon, dust, sea salt, sulfate, and organic carbon).

MERRA-2 provides the concentrations of five externally mixed aerosol species: dust(DUST), sea salt (SS), black carbon (BC), organic carbon (OC), and sulfate (SO4). Following Buchard et al. (2017), we used equation (3) to calculate the surface PM2.5 concentration at each city from 2004 to 2013:

$$\text{PM2.5}_c = \text{DUST}_c + \text{SS}_c + \text{BC}_c + 1.4 \times \text{OC}_c + 0.96 \times \text{SO4}_c \tag{3}$$

4 Results

4.1 Summary Statistics

Table 1. Summary Statistics

	all		farmers		non-farmers	
	Mean	Sd	Mean	Sd	Mean	Sd
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Individual Level Variables						
Age	60.16	8.675	60.68	8.606	58.90	8.714
Hukou	0.220	0.414	0.0663	0.249	0.596	0.491
Gender	0.512	0.500	0.532	0.499	0.463	0.499
Married Status	0.886	0.318	0.879	0.326	0.901	0.298
Years of Schooling	5.661	4.365	4.614	3.993	8.225	4.174
Educational Level	2.876	1.541	2.495	1.323	3.811	1.636
Annual Work Hour in 2011	1,152	1,232	1,150	1,165	1,157	1,385
Annual Work Hour in 2013	1,034	1,197	1,008	1,113	1,098	1,381
Retirement Dummy	0.279	0.449	0.225	0.418	0.413	0.492
Hypertension	0.221	0.415	0.204	0.403	0.263	0.440
Diabetes	0.0487	0.215	0.0372	0.189	0.0769	0.266
Lung Diseases	0.0748	0.263	0.0802	0.272	0.0617	0.241
Heart Diseases	0.0963	0.295	0.0867	0.281	0.120	0.325
Asthma	0.0255	0.158	0.0265	0.161	0.0231	0.150
The Number of Five Chronic Diseases	0.447	0.727	0.407	0.684	0.543	0.812
The Net Wealth of Household	993.8	121,924	-4,492	73,610	14,425	194,295
N	10,904		7,742		3,162	

Notes: The sample is individuals aged from 43 to 78 in 2011 wave and those who answered both 2011 and 2013 waves. Individual characteristics are taken from 2013 wave.

Table 2. Summary Statistics (2)

	All		Farmers		Fon-farmers	
	Mean	Sd	Mean	Sd	Mean	Sd
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
City Level Variables						
The City's Mean PM2.5 (2004-2013)	38.02	11.76	38.51	11.86	36.83	11.41
The Secondary Industry Share of GDP	0.495	0.0792	0.498	0.0759	0.489	0.0863
The Tertiary Industory Share of GDP	0.369	0.0861	0.356	0.0745	0.401	0.103
Mean Wage	44,407	9,247	43,290	7,625	47,141	11,917
Area Size	21,663	37,360	22,273	39,048	20,169	32,823
GDP Per Capita	45,556	38,665	39,591	31,085	60,159	49,897
Population	618.7	422.7	616.9	436.1	622.9	388.1
Population Density	498.0	304.7	491.0	286.1	515.3	345.6
Province Level Variables						
GDP Per Capita	41,732	13,837	40,491	12,362	44,770	16,529
Population	6,421	2,659	6,556	2,594	6,091	2,786
Urban Population Density	2,889	1,228	2,880	1,211	2,910	1,267
The Secondary Industry Share of GDP	0.505	0.0428	0.508	0.0369	0.498	0.0538
The Tertiary Industory Share of GDP	0.389	0.0546	0.383	0.0445	0.402	0.0720
Average Years of Schooling	8.975	0.537	8.904	0.474	9.149	0.635
N	10,904		7,742		3,162	

In China, individuals are classified into either agricultural or non-agricultural hukou, depending on whether they live in rural or urban areas. While most individuals with agricultural hukou live in rural areas, and those with non-agricultural hukou reside in cities, there are exceptions to this rule. In our study, we found that that 78% of individuals lived in rural areas and made a living from agriculture. The male-to-female ratio in our sample was almost 1:1, and most individuals were married, reflecting the fact that our study focused on middle-aged and elderly individuals. For years of education, we categorized education levels into nine categories based on questionnaire responses. Then, we transform them to years of schooling. Notably, the education level was generally low in our sample, consistent with the fact that most individuals were farmers.

In 2011, 4,428 individuals worked fewer than 200 hours per year, while 4,841 individuals worked fewer than 200 hours per year in 2013. This finding may reflect the fact that more individuals retire as they age. One possible reason for the large number of individuals working more than 2,000 hours per year is that many individuals in the study sample were farmers or held both agricultural and non-agricultural jobs.

Table 3. Summary Statistics (3)
the Key Outcome Variables of Those Aged 45-50

	farmers		non-farmers	
	Mean	Sd	Mean	Sd
VARIABLES	(1)	(2)	(3)	(4)
Retirement Dummy	0.0886	0.284	0.124	0.330
Hypertension	0.110	0.313	0.140	0.347
Diabetes	0.0251	0.156	0.0324	0.177
Lung Diseases	0.0418	0.200	0.0280	0.165
Heart Diseases	0.0459	0.209	0.0560	0.230
Asthma	0.0175	0.131	0.00590	0.0766
The Net Wealth of Household	-21,508	121,813	-8,365	192,694
N	1,197		678	

Columns (5)-(6) of Table 1 present the summary statistics for non-farmers. On average, non-farmers are 59 years old, which is slightly higher than the average age of farmers. The sample includes more non-agricultural hukou than agricultural hukou, although the difference is not significant. Non-farmers have a significantly higher level of educational attainment than farmers. Additionally, the retirement rate of non-farmers is 41.3%, which is higher than farmers' retirement rate of 22.5%. However, non-farmers have a higher prevalence of chronic diseases per capita compared to farmers, and their net wealth is also higher.

For the analysis below, we estimate the effect of PM2.5 concentration on retirement of those aged 45 using equation (1). Thus, it is useful to know the summary statistics of the outcome variable of those who are close to age 45. To do so, we pick up individuals whose age is between 45 and 50 and calculate the mean and the standard deviation of the outcome variable. Table 3 shows the mean and the standard deviation of those individuals.

Figure 1 shows the map of average PM 2.5 concentration between 2004 and 2013. Figure 2 shows the histogram of the exposure of the average PM2.5 concentrations of all individuals used in our sample. The histogram shows that the majority of individuals had PM2.5 concentrations

Figure 1: Concentrations of PM2.5 in China from 2004-2013

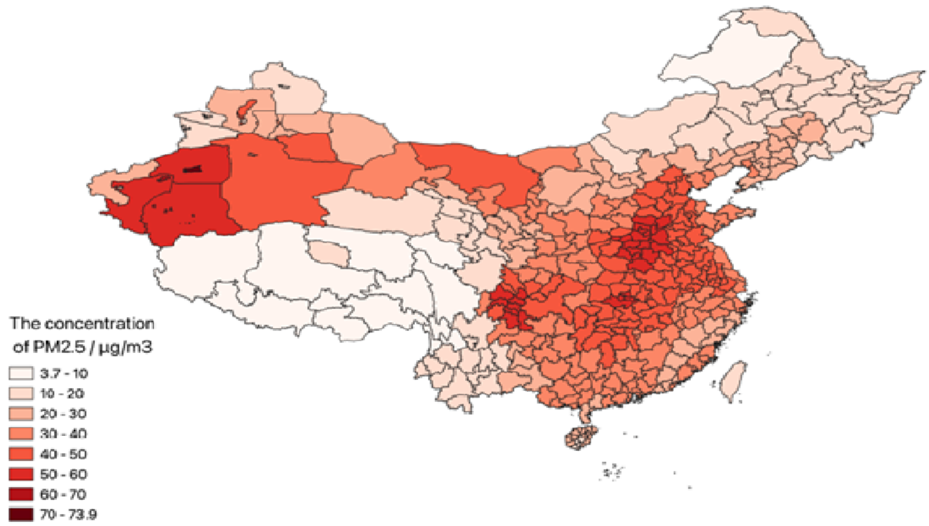
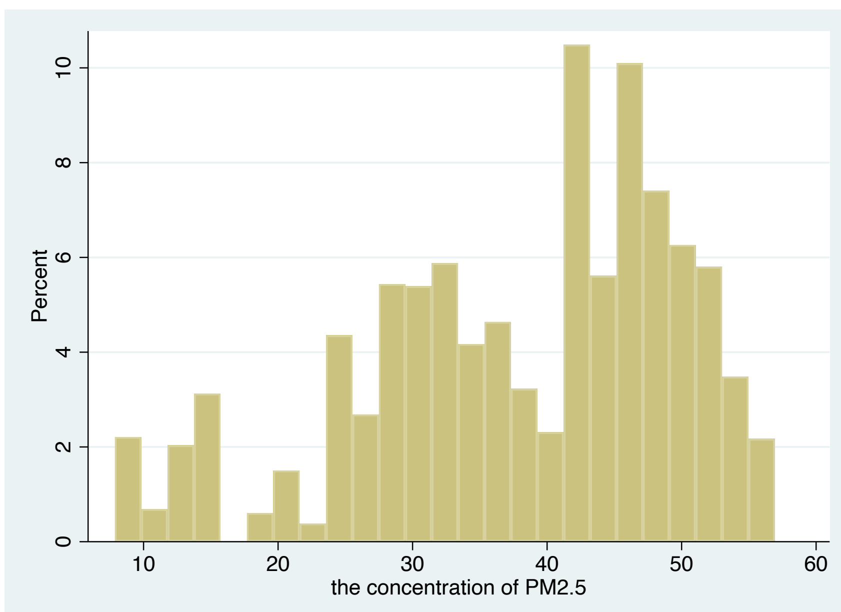


Figure 2: Histogram of PM2.5 Exposure in China



Notes: The unit of the horizontal axis is micrograms per cubic meter. The above histogram is constructed using all individuals used in our regression. Thus, the unit of the histogram is an individual. The above histogram shows that most individuals are exposed to PM2.5 concentrations above 10 micrograms per cubic meter.

above 10 microgram, the WHO guideline value.

4.2 Results of OLS Estimation

Table 4 shows the estimation result of the regression using the retirement dummy as a dependent variable. Panel A shows the results using farmers' sample and Panel B show the the result using non-farmers' sample. Both Panel A and Panel B show that the effect of the retirement is a quadratic function of age. Panel A shows that the estimated coefficient of PM2.5 on the retirement is 0.00093 and it is statistically insignificant. This estimated coefficient implies that a 10 microgram/m^3increase of PM2.5 only increase the probably of the retirement less than 1 percentage point. Thus, the size of the effect of PM2.5 exposure on the retirement for farmers is economically and statistically very small.

Panel B of Table 4 displays the regression results for the retirement of non-farmers. The estimated coefficient of PM2.5 is positive and statistically significant. Panel B shows that the marginal effect of PM2.5 concentrations is a quadratic function of age and highlights the importance of the heterogeneous nonlinear relationship between PM2.5 concentration, age, and retirement. In column (1) of Panel B, we control for individual characteristics such as gender, years of education, hukou, and marital status. The estimated coefficient is 0.00454, indicating that a 10 microgram/m^3increase in PM2.5 concentration increases the probability of retirement by 4.54 percentage points, and the estimated coefficient is statistically significant. In columns (2) and (3), we further include region fixed effects and city characteristics. In column (4), we add province-level covariates. The estimated coefficient is 0.00609, indicating that a 10 microgram/m^3increase in PM2.5 concentration increases the probability of retirement by 6.09 percentage points. Note that for those aged 45-50, the average retirement rate is 12 percent, as shown in Table 3. This implies that a 10 microgram/m^3increase in PM2.5 concentration increases the probability of retirement by about 50 percent. Thus, Panel B of Table 4 shows that PM2.5 has a very large effect on retirement. The estimated regression result indicates that PM2.5 has a more substantial effect on non-farmers than on farmers.

Table 4. The Effect of PM2.5 on Retirement (OLS)

Dependent Variable	Retirement Dummy			
Variables	(1)	(2)	(3)	(4)
Panel A. Farmers' Sample				
PM2.5 – meanPM2.5	0.000979	0.000848	0.00169	0.000931
	(0.000968)	(0.00119)	(0.00158)	(0.00214)
(Age – 45)	-0.000285*	-0.000260	-0.000256	-0.000234
× (PM2.5 – meanPM2.5)	(0.000160)	(0.000164)	(0.000164)	(0.000166)
$(Age – 45)^2$	2.54e-06	1.71e-06	1.53e-06	1.42e-06
× (PM2.5 – meanPM2.5)	(5.17e-06)	(5.08e-06)	(5.05e-06)	(4.98e-06)
Age – 45	-0.00243	-0.00258	-0.00267	-0.00268
	(0.00201)	(0.00205)	(0.00207)	(0.00211)
$(Age – 45)^2$	0.000583***	0.000591***	0.000597***	0.000595***
	(6.22e-05)	(6.25e-05)	(6.26e-05)	(6.30e-05)
R-squared	0.198	0.212	0.214	0.220
Observations	7,742	7,742	7,742	7,742
Panel B. Non-farmers' Sample				
PM2.5 – meanPM2.5	0.00454**	0.00507**	0.00698***	0.00609**
	(0.00187)	(0.00202)	(0.00228)	(0.00264)
(Age – 45)	-0.000721**	-0.000722**	-0.000744**	-0.000771***
× (PM2.5 – meanPM2.5)	(0.000283)	(0.000297)	(0.000288)	(0.000289)
$(Age – 45)^2$	1.83e-05**	1.85e-05**	1.93e-05**	2.03e-05**
× (PM2.5 – meanPM2.5)	(8.38e-06)	(8.87e-06)	(8.51e-06)	(8.53e-06)
Age – 45	0.0332***	0.0332***	0.0329***	0.0328***
	(0.00415)	(0.00420)	(0.00407)	(0.00402)
$(Age – 45)^2$	-0.000135	-0.000147	-0.000144	-0.000148
	(0.000112)	(0.000116)	(0.000112)	(0.000110)
R-squared	0.351	0.358	0.366	0.369
Observations	3,162	3,162	3,162	3,162
Control Variables				
Individual Variables	✓	✓	✓	✓
Region Fixed Effect		✓	✓	✓
Province Variables			✓	✓
City Variables				✓

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5 is the average PM2.5 concentration in the past ten years. meanPM2.5 is the national average of PM2.5 concentration in the past ten years. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. The Effect of PM2.5 on Having Each Disease (OLS): Farmers Sample

Dependent Variables	Hypertension	Diabetes	Lung Diseases	Heart Diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	0.000640	0.000450	0.00261*	0.00296**	0.00113
	(0.00158)	(0.000642)	(0.00140)	(0.00142)	(0.000763)
(Age – 45)	3.67e-05	1.77e-05	-6.36e-05	-0.000188	-1.65e-05
× (PM2.5 – meanPM2.5)	(0.000153)	(6.24e-05)	(0.000151)	(0.000131)	(9.98e-05)
$(\text{Age} - 45)^2$	-2.43e-06	-7.98e-07	3.56e-06	8.38e-06**	1.91e-06
× (PM2.5 – meanPM2.5)	(4.71e-06)	(1.82e-06)	(3.78e-06)	(3.22e-06)	(2.85e-06)
Age – 45	0.0123***	0.00290***	0.00152	0.00724***	-0.000529
	(0.00197)	(0.000755)	(0.00172)	(0.00158)	(0.000947)
$(\text{Age} - 45)^2$	-0.000133**	-6.14e-05***	4.59e-05	-0.000128***	4.37e-05
	(6.12e-05)	(2.27e-05)	(4.62e-05)	(4.06e-05)	(2.76e-05)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
R-squared	0.051	0.011	0.032	0.064	0.015
Observations	7,742	7,742	7,742	7,742	7,742

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5 is the average PM2.5 concentration in the past ten years. meanPM2.5 is the national average of PM2.5 concentration in the past ten years. Column (1)-(5) show different estimation results of different regression which use each type of disease dummy as a dependent variable *** p<0.01, ** p<0.05, * p<0.1.

Tables 5 and Table 6 show the regression results regarding the effect of PM2.5 on the five chronic diseases for the samples of farmers and non-farmers, respectively. We used chronic diseases as the dependent variable to represent the health status of the samples because self-reported health status is always somewhat biased. Previous research has shown that people frequently underestimate their health, particularly when filling out surveys and describing their health negatively (Disney et al., 2006).

The dependent variable in the regression is the dummy variable for having each of the five chronic diseases. Thus, the estimated coefficients show the effect of an increase in PM2.5 concentration on farmers aged 45. Columns (1) to (5) of Table 5 show the effect of PM2.5 on hypertension, diabetes, lung-related diseases, heart-related diseases, and asthma, respectively.

Table 6. The Effect of PM2.5 on Having Each Disease (OLS): Non-Farmers Sample

Dependent Variables	Hypertension	Diabetes	Lung Diseases	Heart Diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	0.00284	0.000242	0.000236	0.00415**	1.70e-05
	(0.00222)	(0.00103)	(0.00137)	(0.00168)	(0.000621)
(Age – 45)	-0.000441	0.000169	1.85e-05	-0.000415**	9.28e-05
× (PM2.5 – meanPM2.5)	(0.000283)	(0.000135)	(0.000177)	(0.000182)	(7.62e-05)
$(\text{Age} - 45)^2$	1.89e-05**	-5.98e-06	9.63e-07	1.38e-05**	-2.94e-06
× (PM2.5 – meanPM2.5)	(8.92e-06)	(4.80e-06)	(6.09e-06)	(6.16e-06)	(2.59e-06)
Age – 45	0.0191***	0.00666***	0.000181	0.00857***	0.00201*
	(0.00333)	(0.00162)	(0.00208)	(0.00200)	(0.00104)
$(\text{Age} - 45)^2$	-0.000272***	-0.000118**	0.000114*	-5.83e-05	-7.80e-06
	(9.57e-05)	(5.59e-05)	(6.85e-05)	(6.60e-05)	(3.44e-05)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
R-squared	0.076	0.028	0.037	0.094	0.022
Observations	3,162	3,162	3,162	3,162	3,162

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5 is the average PM2.5 concentration in the past ten years. meanPM2.5 is the national average of PM2.5 concentration in the past ten years. Column (1)-(5) show different estimation results of different regression which use each type of disease dummy as a dependent variable *** p<0.01, ** p<0.05, * p<0.1.

Table 5 shows that, in terms of the size of the estimated coefficients, the level of PM2.5 concentrations has a relatively large effect on having lung-related diseases and heart-related diseases. For example, column (3) and (4) show that a 10 microgram/m^3 increase in PM2.5 increases the probability of having lung-related and heart-related diseases for farmers aged 45 by 2.61 percentage points and 2.96 percentage points, respectively, and their effects are statistically significant. Note that, in Table 3, the average probability of having lung-related and heart-related diseases for those aged between 45 and 50 is 4.2% and 4.6%, respectively. Thus, a 10 microgram/m^3 increase in PM2.5 increases the probability of having lung-related and heart-related diseases by 62% and 64%, respectively, and these effects are not small.

Table 6 shows the effect of PM 2.5 on having each of the five diseases for the non-farmers in the OLS estimation. Table 6 shows that except for heart-related diseases, the estimated coefficient of PM2.5 is economically very small and statistically insignificant. On the other hand, the effect of PM2.5 on having heart-related diseases is not small. Column (4) of Table 6 shows that 10 microgram/m^3increase of PM2.5 per cubic meter increases the probability of having a heart-related disease by 4.15 percentage points. Table 3 shows that 5.6 percent of non-farmers aged 45-50 have heart-related diseases. Thus, a 10 microgram per cubic meter increase of PM2.5 increases the probability of having heart-related diseases by 74 percent, which is not a small number.

4.3 Results of 2SLS Estimation

In the previous subsection, we used OLS estimation. Note that PM2.5 can be correlated with economic activity, which can affect the retirement decision of workers directly. For example, in an area where economic activity is intensive, the demand for labor can be high, leading people to delay their retirement decision. In such cases, we may underestimate the effect of PM2.5 on retirement. To address this issue in our OLS estimation, we included several variables such as city and province GDP, and GDP shares of specific sectors.

However, it remains possible that certain economic variables may be correlated with both PM2.5 concentration and retirement. To address this potential issue, we use PM2.5 levels from 30 years ago (1980) as an instrumental variable, and we interact it with age and age squared while controlling for current industry structure and economic activities. The key assumption underlying the use of this instrumental variable is that the PM2.5 levels from 30 years ago are not correlated with current retirement behavior, once we account for current economic activities. However, it's worth noting that PM2.5 levels are influenced by various geographic factors such as elevation and proximity to basins, as well as wind conditions. Therefore, current PM2.5 concentrations may still be correlated with past PM2.5 levels even after controlling for current economic conditions.

Table 7 shows the results of the first stage regression in 2SLS estimation for the farmers'

Table 7. First Stage Estimation of 2SLS: Farmers' Sample

Dependent Variable	PM2.5 – meanPM2.5	(Age – 45) × (PM2.5 – meanPM2.5)	$(Age - 45)^2$ × (PM2.5 – meanPM2.5)
Variables	(1)	(2)	(3)
PM2.5(1980)	1.717***	-3.834**	-85.81**
	(0.0719)	(1.626)	(34.24)
(Age – 45) × PM2.5(1980)	-0.000975	2.033***	4.541**
	(0.00101)	(0.0796)	(1.973)
$(Age - 45)^2$ × PM2.5(1980)	6.24e-05*	-0.00389**	1.747***
	(3.33e-05)	(0.00187)	(0.117)
(Age – 45)	0.0213	-36.86***	-125.7***
	(0.0214)	(1.790)	(40.84)
$(Age - 45)^2$	-0.00120*	0.112***	-28.91***
	(0.000687)	(0.0389)	(2.384)
Control Variables			
Individual Variables	✓	✓	✓
Region Fixed Effect	✓	✓	✓
Province Variables	✓	✓	✓
City Variables	✓	✓	✓
F Statatiscs	195.7	439.0	435.5
Observations	7,742	7,742	7,742
R-squared	0.978	0.952	0.929

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5(1980) is PM2.5 concentration in 1980 at the same city minus the national average of PM2.5 concentration in 1980. PM2.5 is the average PM2.5 concentration in the last 10 years (from the year of the survey) in which a respondent lived. meanPM2.5 is the national average of PM2.5 concentration in the last 10 years (from the year of the survey). F statistcs is F value of the null hypothesis that the coefficient of the three nsrumental variables are equal to zero.

sample. In column (1), the dependent variable is the average PM2.5 concentration in a city from 2004 to 2013 where the individual i lived minus the national average of PM2.5 in China from 2004 to 2013. Our primary instrumental variable is the PM2.5 level in 1980 in a city where the respondent lived from 2004 to 2013 minus the national average of PM2.5 in 1980. We also have the interaction of this primary instrumental variable with demeaned age and its square as additional instruments. In column (1), the estimated coefficient of our primary instrument is 1.71 and the standard error is 0.072. Thus, the t-value is more than 23, indicating that our primary instrumental variable works very well.

Table 8. First Stage Estimation of 2SLS: Non-Farmers' Sample

Dependent Variable	PM2.5 – meanPM2.5	(Age – 45) × (PM2.5 – meanPM2.5)	$(Age - 45)^2$ × (PM2.5 – meanPM2.5)
Variables	(1)	(2)	(3)
PM2.5(1980)	1.685***	-4.204**	-90.29**
	(0.0903)	(1.838)	(40.26)
(Age – 45) × PM2.5(1980)	-0.00194	2.008***	1.644
	(0.00253)	(0.117)	(3.513)
$(Age - 45)^2$ × PM2.5(1980)	3.79e-05	-0.00199	1.891***
	(7.71e-05)	(0.00368)	(0.172)
(Age – 45)	0.0206	-35.03***	-26.87
	(0.0501)	(2.394)	(71.37)
$(Age - 45)^2$	-0.000590	0.0327	-33.04***
	(0.00150)	(0.0749)	(3.484)
Control Variables			
Individual Variables	✓	✓	✓
Region Fixed Effect	✓	✓	✓
Province Variables	✓	✓	✓
City Variables	✓	✓	✓
F Statistics	120.2	311.7	280.3
Observations	3,162	3,162	3,162
R-squared	0.967	0.933	0.894

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5(1980) is PM2.5 concentration in 1980 at the same city minus the national average of PM2.5 concentration in 1980. PM2.5 is the average PM2.5 concentration in the last 10 years (from the year of the survey) in which a respondent lived. meanPM2.5 is the national average of PM2.5 concentration in the last 10 years (from the year of the survey). F statistcs is F value of the null hypothesis that the coefficient of the three nsrumental variables are equal to zero. *** p<0.01, ** p<0.05, * p<0.1.

The F-statistic for the null hypothesis that the coefficients of the three instrumental variables are equal to 0 is 195.7, which is higher than the standard threshold of 10. Column (2) estimates the second first stage equation of 2SLS, where the endogenous variable is the interaction of age minus 45 and PM2.5 minus the national average of PM2.5. Again, the F-statistic for the null hypothesis that the coefficients of all three instrumental variables are equal to zero is 439, which is much higher than the standard threshold value. Column (3) estimates the third first stage equation of 2SLS, where the endogenous variable is the interaction of the square of age minus 45, age minus 45, PM2.5 minus the national average of PM2.5. The F-statistic for the null hypothesis that the coefficients of all instruments are equal to zero is 435.5.

Table 9. The Effect of PM2.5 on Retirement (2SLS): the Second Stage Results

Dependent Variable	Retirement Dummy			
Variables	(1)	(2)	(3)	(4)
Panel A. Farmers' Sample				
PM2.5 – meanPM2.5	0.00165	0.00145	0.00195	8.97e-05
	(0.00117)	(0.00123)	(0.00153)	(0.00216)
(Age – 45)	-0.000315**	-0.000286*	-0.000281*	-0.000262
× (PM2.5 – meanPM2.5)	(0.000156)	(0.000158)	(0.000158)	(0.000159)
Age – 45	-0.00232	-0.00250	-0.00254	-0.00251
	(0.00203)	(0.00206)	(0.00207)	(0.00211)
Kleibergen-Paap Rank Wald	153.7	197.9	195.9	211.2
R-squared	0.198	0.211	0.214	0.220
Observations	7,742	7,742	7,742	7,742
Panel B. Non-farmers' Sample				
PM2.5 – meanPM2.5	0.00506***	0.00491**	0.00694***	0.00715**
	(0.00192)	(0.00200)	(0.00228)	(0.00281)
(Age – 45)	-0.000663**	-0.000615*	-0.000631**	-0.000656**
× (PM2.5 – meanPM2.5)	(0.000303)	(0.000314)	(0.000308)	(0.000306)
Age – 45	0.0331***	0.0329***	0.0326***	0.0325***
	(0.00410)	(0.00415)	(0.00400)	(0.00397)
Kleibergen-Paap Rank Wald	99.63	103.2	157.3	137.8
R-squared	0.351	0.358	0.365	0.369
Observations	3,162	3,162	3,162	3,162
Control Variables				
Individual Variables	✓	✓	✓	✓
Region Fixed Effect		✓	✓	✓
Province Variables			✓	✓
City Variables				✓

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5 is the average PM2.5 concentration in the past ten years. meanPM2.5 is the national average of PM2.5 in the past ten years. All specifications above include the square of (Age – 45) and its interaction with (PM2.5 – meanPM2.5). To save the space, we do not provide the estiamted coefficients of those variables, but those coefficients are displayed in Table A1. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8 is the fist stage estimation of 2SLS for non-farmers' sample. In all three cases, the F value of the null hypothesis is above 100, which is higher than the standard threshold value 10.

Panel A of Table 9 shows the results of the second stage of 2SLS estimation for the sample of farmers. The dependent variable is the retirement dummy. The first endogenous explanatory variable is the average PM2.5 concentration in the past 10 years, which is the respondent's exposure to PM2.5 in the past 10 years, minus the national average of PM2.5 in the past 10 years. The second and third endogenous explanatory variables are the interaction of the first endogenous explanatory variable with age minus 45 and the square of age minus 45. In column (1), we include three endogenous explanatory variables, age minus 45, its square, and individual characteristics. In this column, the estimated coefficient of the average PM2.5 in the past 10 years minus the national average of PM2.5 in the past year is 0.00165. Thus, a 10 microgram/m^3increase in PM2.5 increases the probability of retirement of farmers aged 45 by only 1.65 percent. However, the estimated coefficient is not economically or statistically significant.

In column (2) of Panel A of Table 9, we additionally include region fixed effect. The estimated coefficient of the same explanatory variable in column (2) is 0.00145, which is neither economically nor statistically significant. In column (3), the estimated coefficient of the same explanatory variable is 0.00195. In column (4), the estimated coefficient of the same explanatory variable is 0.000089, which is practically equal to zero economically and is not statistically significant.

Regarding the effect of age on the retirement of farmers, Panel A of Table 9 shows that retirement is a quadratic function of age, but the interaction term of age and PM2.5 is close to zero. Thus, overall, we can conclude that the PM2.5 concentration does not significantly affect the retirement decision of farmers.

Panel B of Table 9 shows the effect of PM2.5 concentration on retirement for non-farmers in the second-stage result of our 2SLS estimation. The dependent variable is the retirement dummy. From column (1) to column (4), Panel B shows that PM2.5 concentrations increases the probability of retirement. The interaction term of demeaned age with PM2.5 and the square of demeaned age with PM.25 show statistically and economically significant. This implies that the marginal effect of PM2.5 is a non-linear function of age.

In Panel B of Table 9, column (4) shows that a 10 microgram/m^3increase in PM2.5 increases the probability of retirement by 7.15 percentage points. Note that Table 3 shows that the average probability of retirement for non-farmers aged 45-50 is 12.4 percent. Thus, a 10 microgram/m^3increase in PM2.5 concentration increases the retirement of non-farmers who are 45 years old by 57.6 percent (= $(0.0715/0.124) \times 100$). This implies that PM2.5 concentration has a large effect on retirement.

In 2018, the number of individuals aged 40-49 in China was 228 million. The average share of individuals not in agriculture, fishing, and forestry sectors was 74.2 percent. Assuming this national share applies to those aged 40-49, roughly 169 million individuals are non-farmers in that age group. If China were to reduce the PM2.5 concentration level by 10 micrograms throughout the country, according to column (4) of Panel B of Table 9, 12.1 million people could continue participating in the labor market.

Panel B of Table 9 shows that the interaction term of demeaned PM2.5 and age minus 45 and its square are not zero. This implies that the marginal effect of an increase in PM2.5 depends on age. Figure 3 displays the marginal effect of a 10 microgram/m^3increase in PM2.5 on retirement at different ages in OLS and 2SLS estimations for non-farmers, controlling for all level characteristics. Figure 3 shows that the marginal effects of PM2.5 concentrations are a decreasing function of age in both OLS and 2SLS estimation. Interestingly, the marginal effect of PM2.5 concentration in 2SLS estimation at age 65 becomes almost equal to zero. This suggests that at age 65, most individuals are almost retired regardless of past exposure to PM2.5.

Table 10 shows the effect of PM2.5 concentration on having each of the five chronic diseases for farmers. At first glance, the estimated coefficient of PM2.5 seems not to affect having chronic diseases. However, a closer look shows that the estimated coefficient of PM2.5 in 2SLS estimation is quite similar to that in OLS estimation regarding the effect on lung-related disease. In Table 5, using OLS estimation, the estimated coefficient of PM2.5 concentration on having lung-related and heart-related diseases is 0.00261, and it is statistically significant. In Table 10, using 2SLS estimation, the estimated coefficient of PM2.5 concentration on lung-related diseases is 0.00236. Thus, for lung-related diseases, the estimated coefficients are similar in both OLS and 2SLS estimation. In the case of 2SLS, the standard error becomes larger, reducing the statistical precision of the estimate, but the estimates of OLS and 2SLS are very close. Note that from Table 4, 4.18 percent of farmers aged 45 have lung-related diseases. Thus, Table 10 shows that a 10 microgram per cubic meter increase in PM2.5 increases the probability of having lung-related diseases by 56.4 percent ($= (0.0236/0.0418) \times 100$).

Table 11 shows the effect of PM2.5 concentration on having each of the five chronic diseases for non-farmers. Table 11 shows that a higher PM2.5 concentration increases the likelihood of having

Table 10. The Effect of PM2.5 on Having Each Disease (2SLS)
The Second Stage Estimation: Farmers' Sample

Dependent Variables	Hypertension	Diabetes	Lung Diseases	Heart Diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	-0.000530	0.000220	0.00236	0.00174	0.00112
	(0.00162)	(0.000685)	(0.00161)	(0.00142)	(0.000885)
(Age – 45)	2.67e-05	1.27e-05	-7.50e-06	-0.000112	-3.22e-05
× (PM2.5 – meanPM2.5)	(0.000161)	(7.40e-05)	(0.000167)	(0.000146)	(0.000114)
Age – 45	0.0124***	0.00294***	0.00124	0.00689***	-0.000445
	(0.00201)	(0.000787)	(0.00178)	(0.00164)	(0.000941)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
Kleibergen-Paap Rank Wald	211.2	211.2	211.2	211.2	211.2
R-squared	0.051	0.011	0.032	0.064	0.015
Observations	7,742	7,742	7,742	7,742	7,742

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5 is the average PM2.5 concentration in the past ten years.meanPM2.5 is the national average of PM2.5 in the past ten years. Column (1)-(5) show different estimation results of different regressions which use each type of disease dummy as a dependent variable. All specifications above include the square of (Age – 45) and its interaction with (PM2.5 – meanPM2.5). To save the space, we do not provide the estiamted coefficients of those variables, but those coefficients are displayed in Table A2. *** p<0.01, ** p<0.05, * p<0.1.

heart-related diseases. Column (4) shows that a 10 microgram/m^{3}increase of PM2.5 increases the likelihood of having heart-related diseases by 3.25 percentage points. The estimated coefficients are similar in both OLS and 2SLS. In OLS, the effect on having heart-related diseases is 0.00415 and in 2SLS it is 0.00325. Note that, for non-farmers, the average probability of heart-related diseases is 5.6 percent. Thus, a 10 microgram/m^{3}increase of PM2.5 concentration increases the likelihood of having heart-related diseases by 58 percent.

Farmers tend to suffer from lung-related diseases. However, Table 9 shows that the retirement behavior of farmers is not affected by a higher PM2.5 level, while the retirement behavior of non-farmers is affected by it. One natural question would be why we observe such different behavior regarding the response to PM2.5. Our conjecture is that since non-farmers have more wealth, they can choose to retire early when their health becomes bad. In fact, Table 2 shows that

Table 11. The Effect of PM2.5 on Having Each Disease (2SLS)
The Second Stage Estimatoin: Non-farmers' Sample

Dependent Variables	Hypertension	Diabetes	Lung Diseases	Heart Diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	0.00146	0.000795	0.000349	0.00325*	-0.000521
	(0.00272)	(0.00123)	(0.00151)	(0.00187)	(0.000715)
(Age – 45)	-0.000504	0.000171	7.80e-07	-0.000392**	0.000145
× (PM2.5 – meanPM2.5)	(0.000307)	(0.000164)	(0.000189)	(0.000184)	(9.37e-05)
Age – 45	0.0194***	0.00666***	0.000230	0.00851***	0.00184*
	(0.00331)	(0.00158)	(0.00208)	(0.00197)	(0.00106)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
Kleibergen-Paap Rank Wald	137.8	137.8	137.8	137.8	137.8
R-squared	0.076	0.028	0.037	0.094	0.022
Observations	3,162	3,162	3,162	3,162	3,162

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5 is the average PM2.5 concentration in the past ten years. meanPM2.5 is the national average of PM2.5 in the past ten years. Column (1)-(5) show different estimation results of different regressions which use each type of disease dummy as a dependent variable. All specifications above include the square of (Age – 45) and its interaction with (PM2.5 – meanPM2.5). To save the space, we do not provide the estiamted coefficients of those variables, but those coefficients are displayed in Table A3. *** p<0.01, ** p<0.05, * p<0.1.

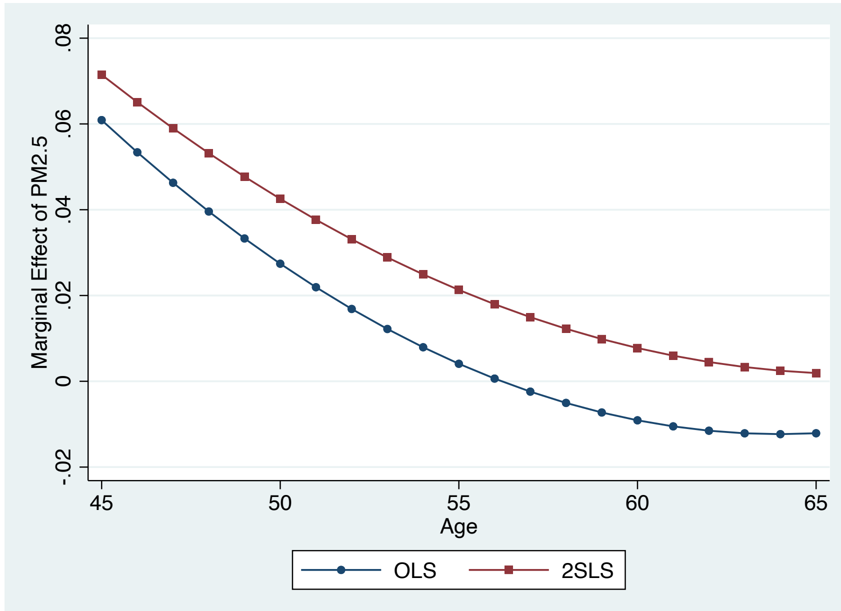
farmers' net wealth (the balance of assets minus debt) is negative, while non-farmers' net wealth is positive. Thus, our interpretation is consistent with the summary statistics.

4.4 A Potential Mechanism

As we discussed in the introduction, a higher PM2.5 level can affect the retirement decision through two channels. The first channel is the disutility channel, where increased air pollution leads to deteriorating health among workers, causing disutility of work to increase and inducing early retirement. The second channel is the productivity channel, where reduced health lowers the productivity of workers, decreasing their financial wealth. Since leisure is a normal good, a lower wealth level could postpone a decision to retire.

To explore policy implications, it is important to identify the mechanism behind the outcome.

Figure 3: Marginal Effects of a 10 Microgram/m^3 Increase of PM 2.5 on Retirement Probability at Each Age



Notes: Each point denotes the marginal effect of 10 microgram/m^3increase of PM2.5 on the probability of retirement. A blue line with circles denote the marginal effect based on OLS estimation and a red line with square denotes the marginal effect based on 2SLS estimation.

However, since the outcome is a combination of the two channels, it is difficult to infer which channel is at work. To gain insight, we regressed financial wealth (the value of assets minus debt) on PM2.5. We used Equation (1) and set net wealth as the outcome variable. Table 12 shows the estimation results, with columns (1) and (2) displaying the results using the farmers' sample, and columns (3) and (4) displaying the results using the non-farmers' sample

Columns (1) and (2) of Table 12 use OLS and 2SLS estimation for farmers' sample, respectively. Both columns show that higher PM2.5 levels decrease net wealth, but the coefficient is very small. For instance, column (2) demonstrates that a 10 microgram/m^3increase of PM2.5 concentration decreases net wealth by 0.11 Chinese yuan. However, the net wealth is already -4000 yuan, indicating that the estimated coefficient is negligible when compared to the average PM2.5 value. Similarly, for non-farmers, both OLS and 2SLS show that the estimated coefficient of PM2.5 is positive but statistically insignificant, with a value of 0.0327. This implies that a 10 microgram/m^3increase of PM2.5 concentration increases net wealth by only 0.33 Chinese yuan,

Table 12. The Effect of PM2.5 on Net Wealth (OLS and 2SLS)

Dependent Variables	Net Wealth			
Sample	Farmers		Non-Farmers	
Estimation Method	OLS	2SLS	OLS	2SLS
Variables	(1)	(2)	(3)	(4)
PM2.5 – meanPM2.5	-0.0104	-0.000649	0.0336	0.0327
	(0.0130)	(0.0144)	(0.0230)	(0.0242)
(Age – 45)	0.00148	0.000793	0.000830	0.00170
× (PM2.5 – meanPM2.5)	(0.00131)	(0.00145)	(0.00167)	(0.00190)
$(Age – 45)^2$	-4.03e-05	-2.05e-05	-1.74e-05	-4.48e-05
× (PM2.5 – meanPM2.5)	(3.85e-05)	(4.34e-05)	(5.16e-05)	(6.13e-05)
Age – 45	-0.0642***	-0.0611***	-0.0314	-0.0343
	(0.0165)	(0.0173)	(0.0254)	(0.0247)
$(Age – 45)^2$	-0.000931*	-0.00102**	-0.000172	-8.32e-05
	(0.000472)	(0.000494)	(0.000764)	(0.000740)
Control Variables				
Individual Variables	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓
City Variables	✓	✓	✓	✓
Kleibergen-Paap Rank Wald		211.2		137.8
R-squared	0.209	0.209	0.100	0.100
Observations	7,742	7,742	3,162	3,162

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each city. PM2.5 is the average PM2.5 concentration in the past ten years. meanPM2.5 is the national average of PM2.5 concentration in the past ten years. *** p<0.01, ** p<0.05, * p<0.1.

which is too small to be distinguished from zero.

Therefore, Table 12 shows that the effect of PM2.5 concentration on net wealth is practically zero, implying that the main channel through which PM2.5 concentration affects retirement is the disutility channel through deteriorating health.

4.5 Potential Bias due to Internal Migration

One potential concern regarding our estimation results is the possibility of bias since a respondent's current address may differ from where they lived in the past 10 years, potentially resulting in a downward bias in our estimates. For example, a respondent may have recently relocated to their current city due to asthma caused by air pollution in their previous city.

Table 13. The Effect of PM2.5 on Retirement : Robustness Checks
Using Province Level PM2.5 for Analysis : the Second Stage Results of 2SLS

Dependent Variables	Retirement Dummy			
Variables	(1)	(2)	(3)	(4)
Panel A. Farmers' Sample				
PM2.5 – meanPM2.5	0.00132	-0.00116	-0.000409	-0.00100
	(0.00204)	(0.00272)	(0.00368)	(0.00415)
(Age – 45)	-0.000151	-0.000131	-0.000133	-9.98e-05
× (PM2.5 – meanPM2.5)	(0.000318)	(0.000299)	(0.000302)	(0.000300)
Age – 45	-0.00365	-0.00344	-0.00345	-0.00348
	(0.00224)	(0.00221)	(0.00220)	(0.00222)
Kleibergen-Paap Rank Wald	50.40	54.79	54.31	72.32
R-squared	0.192	0.210	0.212	0.219
Observations	7,742	7,742	7,742	7,742
Panel B. Non-farmers' Sample				
PM2.5 – meanPM2.5	0.00635**	0.00459	0.00834**	0.00783*
	(0.00250)	(0.00287)	(0.00350)	(0.00462)
(Age – 45)	-0.000862**	-0.000808*	-0.000818*	-0.000858**
× (PM2.5 – meanPM2.5)	(0.000423)	(0.000442)	(0.000446)	(0.000428)
Age – 45	0.0327***	0.0325***	0.0321***	0.0321***
	(0.00324)	(0.00346)	(0.00329)	(0.00315)
Kleibergen-Paap Rank Wald	45.97	50.50	29.24	22.21
R-squared	0.351	0.358	0.366	0.369
Observations	3,162	3,162	3,162	3,162
Control Variables				
Individual Variables	✓	✓	✓	✓
Region Fixed Effect		✓	✓	✓
Province Variables			✓	✓
City Variables				✓

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each province. PM2.5 is the average PM2.5 concentration in the past ten years in each province. meanPM2.5 is the national average of PM2.5 in the past ten years. All specifications above include the square of (Age – 45) and its interaction with (PM2.5 – meanPM2.5). To save the space, we do not provide the estiamted coefficients of those variables, but those coefficients are displayed in Table A4. *** p<0.01, ** p<0.05, * p<0.1.

However, we argue that this possibility is likely to be very small. In our dataset, among 7,742 farmers, 99.01% of farmers (7,666 farmers) have been living in the current place since at least 2004, and only 76 farmers moved to the current place after 2004. Similarly, among 3,162 non-farmers, 96.52% of non-farmers (3,045 non-farmers) have been living in the current place since at least 2004, and only 110 non-farmers moved to the current address after 2004. Thus, the bias due to internal migration is likely to be small.

To further check the robustness of our estimation, we conducted a 2SLS estimation using province-level PM2.5 concentration instead of city-level PM2.5 as the key explanatory variable. We also used the province-level PM2.5 30 years ago as the instrumental variable, assuming that respondents are likely to have moved within the same province when they relocated due to air pollution. Moving across provinces involves higher psychological and search costs of finding a new job, so individuals are more likely to move within the same province.

Table 13 presents the results of the 2SLS estimation of the effect of PM2.5 on retirement, using province-level PM2.5 levels as the endogenous explanatory variable instead of city-level PM2.5 levels. The instrumental variable is also the province-level PM2.5 from 30 years ago. The estimation results of 2SLS using province-level PM2.5 are quite similar to those using city-level PM2.5, as shown in Table 9. For instance, in column 4 of Panel B of Table 13, the estimated coefficient of demeaned PM2.5 is 0.00783, while in column 4 of Panel B of Table 9, the estimated coefficient of demeaned PM2.5 is 0.00715.

Appendix A includes Table A5 and Table A6, which present the 2SLS estimation results of the effect of PM2.5 concentration on different types of diseases for farmers (Table A5) and non-farmers (Table A6), using province-level PM2.5 concentration. The estimation results in Table A5 and Table A6 are quite similar to those in Table 10 and Table 11. These tables suggest that the bias due to internal migration is likely to be very small.

However, one might argue that it is possible that those 186 respondents(76 farmers and 110 non-farmers) might have moved across provinces and those internal migration causes bias in our estimates. To check such a possibility, we drop those 186 respondents, estimate our model using the city level and compare the estimates with the estimates in Table 9. Table A7 shows the result of 2SLS regression dropping 186 respondents. The key explanatory variable is the city level demeaned PM2.5 as in Table 9. The estimated coefficient of Table A5 is quite similar to the estimated coefficients of Table 9. For example, in column (4) of Panel B of Table A7, the estimated coefficient of the demeaned PM2.5 is 0.00656 while in column (4) of Panel B of Table 9, the estimated coefficient of the same variable is 0.00715. Thus, those two estimated coefficients are quite similar.

In summary, Table 13, Table A5, Table A6 and Table A7 indicate that the bias due to the internal migration is not likely to affect our conclusion.

4.6 Other Potential Bias

One criticism of our estimation strategy using both OLS and 2SLS is that the concentration of PM2.5 may be correlated with unobserved industry characteristics in each city, which could also be related to the retirement patterns of the city. As a result, our estimates of the effect of PM2.5 may include not only the direct effect of PM2.5 concentration but also the indirect effect of these unobserved industry characteristics. Since we only have cross-sectional data, it is difficult to directly address this issue. However, we believe that if such a bias exists, it is likely to result in an underestimation of the effect of PM2.5.

In both OLS and 2SLS, as we move from column (1) to column (4) in Tables 4 and 9, we include more province and city characteristics, which implies that we control for more industry characteristics. However, we observe that the estimated coefficient of PM2.5 in column (1) is smaller than the estimated coefficient in column (4). If the estimated coefficient of PM2.5 were mainly driven by unobserved industry characteristics of each city, then the estimated coefficients should decrease as we include more province and city characteristics as control variables. Therefore, this suggests that even if there is an unobserved industry effect, it is likely that our estimated coefficients of PM2.5 are underestimated rather than overestimated due to this bias.

5 Discussion and Conclusion

It is well-known that PM2.5 particles are very small and can penetrate deep into the human body beyond the lungs, entering the bloodstream and causing inflammation in various organs. Medical literature has indicated that exposure to PM2.5 can lead to heart diseases, high blood pressure, and other illnesses.

Previous studies on retirement behavior among middle-aged and older adults have mainly focused on analyzing the effects of factors such as health status, financial status, and retirement preparation on retirement. However, in our study, we aimed to investigate the impact of PM2.5

concentration on retirement behavior among middle-aged and older adults in China. This is particularly crucial because China is known for its severe air pollution. Although there are many studies that examine the effect of air pollution on labor supply, to the best of our knowledge, this is the first paper to study the effect of air pollution on the retirement decision.

To achieve this, we divided the study sample into two groups: farmers and non-farmers. We separately examined the effects of PM2.5 on their retirement behavior and health status.

Our study indicates that higher levels of PM2.5 concentration significantly increase the probability of retirement among non-farmers. Specifically, a 10 microgram/m^3increase in PM2.5 concentration (which is roughly equivalent to one standard deviation of PM2.5 concentrations) raises the retirement probability of non-farmers aged 45 by 7.15 percentage points. Since the retirement rate for non-farmers aged 45-50 is 12.4 percent, this means that a 10 microgram/m^3increase in PM2.5 concentration increases the retirement probability of non-farmers aged 45 by 57.6 percent. In contrast, we did not find any evidence that higher PM2.5 concentrations increase the probability of early retirement among farmers. The estimated coefficients for farmers are statistically and economically insignificant. We also observed that large heterogeneous effect of PM 2.5 concentrations on the retirement behavior. As Figure 4 shows, the marginal effect of an increase of PM2.5 concentration highly depends on age. At age 65, the marginal effect becomes almost equal to zero.

Regarding the effect of PM2.5 concentration on diseases, we found that for non-farmers aged 45, a 10 microgram/m^3increase in PM2.5 concentration raises the probability of heart-related diseases by 58 percent. For farmers, a 10 microgram/m^3increase in PM2.5 concentration increases the probability of both lung-related and heart-related diseases by 56 percent.

As discussed in the introduction, air pollution affects the retirement behavior of middle and elderly people through two mechanisms. The first mechanism is by decreasing productivity and financial wealth. A lower wage decreases the financial wealth of middle and elderly people, and since leisure is a normal good, lower wealth due to air pollution is likely to decrease the probability of early retirement. The second mechanism is by increasing the disutility of the labor supply due to bad health caused by air pollution. Higher disutility due to air pollution is likely to increase the probability of early retirement. Our regression on net wealth shows that air pollution does not decrease the net wealth of farmers and non-farmers, suggesting that the main mechanism through which a higher PM2.5 concentration increases the probability of retirement for non-farmers is likely the disutility channel.

Overall, our results are consistent with previous studies. Several recent studies have examined the effects of PM 2.5 on economic outcomes such as migration, obesity, and labor supply (Deschenes et al., 2020; Huang et al., 2021; Li and Li, 2022; Chen et al., 2022; Wu et al., 2023). For example, Huang et al. (2021) show that 10 microgram/m^3decrease of PM2.5 increase labor supply by 37.5 percent. Wu et al. (2023) shows that one percent increase of PM2.5 decrease the ratio of employment to the whole population by 0.02018 unit. This can be translated into that 10 microgram/m^3increase of PM2.5 concentration increase the employment to population ratio 45 percentage points. Additionally, studies in environmental science and health science have reported that higher air pollution in China affects the health status of Chinese people (Chen et al., 2013; Pui et al., 2014; Cakmak et al., 2007; Xue et al., 2018).

As the aging population in China continues to increase, the retirement decisions of Chinese individuals have a significant impact on the economy. In 2018, the number of people aged 40-49 in China was 228 million, with an average of 74.2% working in non-agricultural, fishing, and forestry sectors. Assuming this national average applies to the 40-49 age group, approximately 169 million individuals in this group are non-farmers. Based on the findings in column (4) of Panel B in Table 9, reducing PM2.5 concentration levels by 10 micrograms across China could enable 12.1 million people to remain in the labor market.

China, South Asia, and Central Africa are among the most polluted regions in the world, with high population densities. Our study suggests that in these areas, many middle-aged and elderly individuals may have left the labor market due to air pollution. Therefore, future studies should focus on quantifying the impact of air pollution on retirement decisions in developing countries such as South Asia and Africa.

References

Andersen, Zorana J, Ole Raaschou-Nielsen, Matthias Ketzel, Steen S Jensen, Martin Hvidberg, Steffen Loft, Anne Tjønneland, Kim Overvad, and Mette Sørensen (2012) "Diabetes incidence and long-term exposure to air pollution: a cohort study," *Diabetes care*, Vol. 35, pp. 92–98.

Aragon, Fernando M, Juan Jose Miranda, and Paulina Oliva (2017) "Particulate matter and labor supply: The role of caregiving and non-linearities," *Journal of Environmental Economics and Management*, Vol. 86, pp. 295–309.

Au, Doreen Wing Han, Thomas F Crossley, and Martin Schellhorn (2005) "The effect of health changes and long-term health on the work activity of older Canadians," *Health economics*, Vol. 14, pp. 999–1018.

Bell, Michelle L, Keita Ebisu, Brian P Leaderer, Janneane F Gent, Hyung Joo Lee, Petros Koutrakis, Yun Wang, Francesca Dominici, and Roger D Peng (2014) "Associations of PM2. 5 constituents and sources with hospital admissions: analysis of four counties in Connecticut and Massachusetts (USA) for persons greater than 65 years of age," *Environmental health perspectives*, Vol. 122, pp. 138–144.

Bender, Keith A (2012) "An analysis of well-being in retirement: The role of pensions, health, and 'voluntariness' of retirement," *The Journal of Socio-Economics*, Vol. 41, pp. 424–433.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French (2023) "The impact of health on labor supply near retirement," *Journal of Human Resources*, Vol. 58, pp. 282–334.

Bound, John, Michael Schoenbaum, Todd R Stinebrickner, and Timothy Waidmann (1999) "The dynamic effects of health on the labor force transitions of older workers," *Labour Economics*, Vol. 6, pp. 179–202.

Bound, John, Todd Stinebrickner, and Timothy Waidmann (2010) "Health, economic resources and the work decisions of older men," *Journal of econometrics*, Vol. 156, pp. 106–129.

Buchard, V, CA Randles, AM Da Silva, A Darmenov, PR Colarco, R Govindaraju, R Ferrare, J Hair, AJ Beyersdorf, LD Ziemba et al. (2017) "The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies," *Journal of Climate*, Vol. 30, pp. 6851–6872.

Cakmak, Sabit, Robert E Dales, and Claudia Blanco Vidal (2007) "Air pollution and mortality in Chile: susceptibility among the elderly," *Environmental Health Perspectives*, Vol. 115, pp. 524–527.

Chen, Shuai, Paulina Oliva, and Peng Zhang (2022) "The effect of air pollution on migration: evidence from China," *Journal of Development Economics*, Vol. 156, p. 102833.

Chen, Yuyu, Avraham Ebenstein, Michael Greenstone, and Hongbin Li (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*, Vol. 110, pp. 12936–12941.

Clarke, Philippa, Victor W Marshall, and David Weir (2012) "Unexpected retirement from full time work after age 62: consequences for life satisfaction in older Americans," *European Journal of Ageing*, Vol. 9, pp. 207–219.

De Wind, Astrid, Goedele A Geuskens, Kerstin G Reeuwijk, Marjan J Westerman, Jan Fekke Ybema, Alex Burdorf, Paulien M Bongers, and Allard J Van der Beek (2013) "Pathways through which health influences early retirement: a qualitative study," *BMC Public Health*, Vol. 13, pp. 1–9.

Deschenes, Olivier, Huixia Wang, Si Wang, and Peng Zhang (2020) "The effect of air pollution on body weight and obesity: Evidence from China," *Journal of Development Economics*, Vol. 145, p. 102461.

Disney, Richard, Carl Emmerson, and Matthew Wakefield (2006) "Health and Retirement in Britain: A panel data-based analysis," *Journal of health economics*, Vol. 25, pp. 621–649.

Du, Yixing, Xiaohan Xu, Ming Chu, Yan Guo, and Junhong Wang (2016) "Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence," *Journal of thoracic disease*, Vol. 8, p. E8.

Dwyer, Debra Sabatini and Olivia S Mitchell (1999) "Health problems as determinants of retirement: Are self-rated measures endogenous?" *Journal of health economics*, Vol. 18, pp. 173–193.

Fang, Evandro Fei, Morten Scheibye-Knudsen, Heiko J Jahn, Juan Li, Li Ling, Hongwei Guo, Xinqiang Zhu, Victor Preedy, Huiming Lu, Vilhelm A Bohr et al. (2015) "A research agenda for aging in China in the 21st century," *Ageing research reviews*, Vol. 24, pp. 197–205.

Franklin, Barry A, Robert Brook, and C Arden Pope III (2015) "Air pollution and cardiovascular disease," *Current problems in cardiology*, Vol. 40, pp. 207–238.

Gu, Hejun, Fengqin Bian, and Ehsan Elahi (2020) "Effect of air pollution on female labor supply: An empirical analysis based on data of labor force dynamic survey of China," *Social Work in Public Health*, Vol. 35, pp. 187–196.

Hanna, Rema and Paulina Oliva (2015) "The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City," *Journal of Public Economics*, Vol. 122, pp. 68–79.

Hausman, Jerry A, Bart D Ostro, and David A Wise (1984) "Air pollution and lost work."

He, Jiaxiu, Haoming Liu, and Alberto Salvo (2019) "Severe air pollution and labor productivity: Evidence from industrial towns in China," *American Economic Journal: Applied Economics*, Vol. 11, pp. 173–201.

Huang, Qiaolong, Yu Yvette Zhang, Qin Chen, and Manxiu Ning (2021) "Does Air Pollution Decrease Labor Supply of the Rural Middle-Aged and Elderly?" *Sustainability*, Vol. 13, p. 2906.

Ilmakunnas, Pekka and Seija Ilmakunnas (2018) "Health and retirement age: comparison of expectations and actual retirement," *Scandinavian journal of public health*, Vol. 46, pp. 18–31.

Jones, Andrew M, Nigel Rice, and Jennifer Roberts (2010) "Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS," *Economic Modelling*, Vol. 27, pp. 866–880.

Kan, Haidong, Stephanie J London, Guohai Chen, Yunhui Zhang, Guixiang Song, Naiqing Zhao, Lili Jiang, and Bingheng Chen (2008) "Season, sex, age, and education as modifiers of the effects of outdoor air pollution on daily mortality in Shanghai, China: The Public Health and Air Pollution in Asia (PAPA) Study," *Environmental health perspectives*, Vol. 116, pp. 1183–1188.

Krämer, Ursula, Christian Herder, Dorothea Sugiri, Klaus Strassburger, Tamara Schikowski, Ulrich Ranft, and Wolfgang Rathmann (2010) "Traffic-related air pollution and incident type 2 diabetes: results from the SALIA cohort study," *Environmental health perspectives*, Vol. 118, pp. 1273–1279.

Li, Xiaoqin and Yonghui Li (2022) "The impact of perceived air pollution on labour supply: Evidence from China," *Journal of Environmental Management*, Vol. 306, p. 114455.

Liang, Ruijuan, Biao Zhang, Xiaoyi Zhao, Yanping Ruan, Hui Lian, and Zhongjie Fan (2014) "Effect of exposure to PM2. 5 on blood pressure: a systematic review and meta-analysis," *Journal of hypertension*, Vol. 32, pp. 2130–2141.

Lin, Hualiang, Yanfei Guo, Yang Zheng, Qian Di, Tao Liu, Jianpeng Xiao, Xing Li, Weilin Zeng, Lenise A Cummings-Vaughn, Steven W Howard et al. (2017) "Long-term effects of ambient PM2. 5 on hypertension and blood pressure and attributable risk among older Chinese adults," *Hypertension*, Vol. 69, pp. 806–812.

McGarry, Kathleen (2004) "Health and retirement: do changes in health affect retirement expectations?" *Journal of Human Resources*, Vol. 39, pp. 624–648.

National Bureau of Statistics of China (2011) "The 6th national population census."

——— (2021) "The 7th national population census."

Ostro, Bart D (1983) "The effects of air pollution on work loss and morbidity," *Journal of Environmental Economics and Management*, Vol. 10, pp. 371–382.

Pönkä, Antti (1990) "Absenteeism and respiratory disease among children and adults in Helsinki in relation to low-level air pollution and temperature," *Environmental Research*, Vol. 52, pp. 34–46.

Pui, David YH, Sheng-Chieh Chen, and Zhili Zuo (2014) "PM2. 5 in China: Measurements, sources, visibility and health effects, and mitigation," *Particuology*, Vol. 13, pp. 1–26.

Sun, Qinghua, Peibin Yue, Zhekang Ying, Arturo J Cardounel, Robert D Brook, Robert Devlin, Jing-Shiang Hwang, Jay L Zweier, Lung Chi Chen, and Sanjay Rajagopalan (2008) "Air pollution exposure potentiates hypertension through reactive oxygen species-mediated activation of Rho/ROCK," *Arteriosclerosis, thrombosis, and vascular biology*, Vol. 28, pp. 1760–1766.

World Health Organization and European Centre for Environment (2021) *WHO global air quality guidelines: particulate matter (PM2. 5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*: World Health Organization.

Wu, Beibei, Tiemei Yan, and Ehsan Elahi (2023) "The impact of environmental pollution on labor supply: empirical evidence from China," *Environmental Science and Pollution Research*, Vol. 30, pp. 25764–25772.

Xue, Xiaoxia, Jianping Chen, Baijun Sun, Baosen Zhou, and Xuelian Li (2018) "Temporal trends in respiratory mortality and short-term effects of air pollutants in Shenyang, China," *Environmental Science and Pollution Research*, Vol. 25, pp. 11468–11479.

Ying, Zhekang, Xiaohua Xu, Yuntao Bai, Jixin Zhong, Minjie Chen, Yijia Liang, Jinzhuo Zhao, Dongyao Liu, Masako Morishita, Qinghua Sun et al. (2014) "Long-term exposure to concentrated ambient PM2. 5 increases mouse blood pressure through abnormal activation of the sympathetic nervous system: a role for hypothalamic inflammation," *Environmental health perspectives*, Vol. 122, pp. 79–86.

Zhao, Yaohui, John Strauss, Xinxin Chen, Yafeng Wang, Jinquan Gong, Qinqin Meng, Gewei Wang, and Huali Wang (2020) "China health and retirement longitudinal study wave 4 user's guide," *National School of Development, Peking University*, pp. 5–6.

Zhao, Yaohui, John Strauss, Gonghuan Yang, John Giles, Peifeng Hu, Yisong Hu, Xiaoyan Lei, Albert Park, James P Smith, and Yafeng Wang (2013) "China health and retirement longitudinal study–2011–2012 national baseline users' guide," *Beijing: National School of Development, Peking University*, Vol. 2.

Zhou, Xinping, Yangyang Xu, Shuo Yuan, Cai Wu, and Hao Zhang (2015) "Performance and potential of solar updraft tower used as an effective measure to alleviate Chinese urban haze problem," *Renewable and Sustainable Energy Reviews*, Vol. 51, pp. 1499–1508.

Appendicies

Table A1. The Effect of PM2.5 on Retirement (2SLS): the Second Stage Results

Dependent Variable	Retirement Dummy			
Variables	(1)	(2)	(3)	(4)
Panel A. Farmers' Sample				
PM2.5 – meanPM2.5	0.00165	0.00145	0.00195	8.97e-05
	(0.00117)	(0.00123)	(0.00153)	(0.00216)
(Age – 45)	-0.000315**	-0.000286*	-0.000281*	-0.000262
× (PM2.5 – meanPM2.5)	(0.000156)	(0.000158)	(0.000158)	(0.000159)
$(Age – 45)^2$	2.08e-06	1.30e-06	1.09e-06	1.29e-06
× (PM2.5 – meanPM2.5)	(5.16e-06)	(4.98e-06)	(4.97e-06)	(4.88e-06)
Age – 45	-0.00232	-0.00250	-0.00254	-0.00251
	(0.00203)	(0.00206)	(0.00207)	(0.00211)
$(Age – 45)^2$	0.000586***	0.000594***	0.000600***	0.000595***
	(6.25e-05)	(6.28e-05)	(6.28e-05)	(6.33e-05)
Kleibergen-Paap Rank Wald	153.7	197.9	195.9	211.2
R-squared	0.198	0.211	0.214	0.220
Observations	7,742	7,742	7,742	7,742
Panel B. Non-farmers' Sample				
PM2.5 – meanPM2.5	0.00506***	0.00491**	0.00694***	0.00715**
	(0.00192)	(0.00200)	(0.00228)	(0.00281)
(Age – 45)	-0.000663**	-0.000615*	-0.000631**	-0.000656**
× (PM2.5 – meanPM2.5)	(0.000303)	(0.000314)	(0.000308)	(0.000306)
$(Age – 45)^2$	1.57e-05*	1.38e-05	1.45e-05	1.54e-05*
× (PM2.5 – meanPM2.5)	(9.17e-06)	(9.52e-06)	(9.30e-06)	(9.23e-06)
Age – 45	0.0331***	0.0329***	0.0326***	0.0325***
	(0.00410)	(0.00415)	(0.00400)	(0.00397)
$(Age – 45)^2$	-0.000128	-0.000132	-0.000129	-0.000132
	(0.000112)	(0.000115)	(0.000110)	(0.000109)
Kleibergen-Paap Rank Wald	99.63	103.2	157.3	137.8
R-squared	0.351	0.358	0.365	0.369
Observations	3,162	3,162	3,162	3,162
Control Variables				
Individual Variables	✓	✓	✓	✓
Region Fixed Effect		✓	✓	✓
Province Variables			✓	✓
City Variables				✓

Notes: This is the detail version of Table 9.It shows the estimated coefficients of other variables of the specifications presented in Table 9. Notes of Table 9 apply. *** p<0.01, ** p<0.05, * p<0.1.

Table A2. The Effect of PM2.5 on Having Each Disease (2SLS)
The Second Stage Estimation: Farmers' Sample

Dependent Variables	Hypertension	Diabetes	Lung diseases	Heart diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	-0.000530	0.000220	0.00236	0.00174	0.00112
	(0.00162)	(0.000685)	(0.00161)	(0.00142)	(0.000885)
(Age – 45)	2.67e-05	1.27e-05	-7.50e-06	-0.000112	-3.22e-05
× (PM2.5 – meanPM2.5)	(0.000161)	(7.40e-05)	(0.000167)	(0.000146)	(0.000114)
$(\text{Age} - 45)^2$	-2.10e-06	-3.96e-07	2.69e-06	5.61e-06	2.78e-06
× (PM2.5 – meanPM2.5)	(5.02e-06)	(2.10e-06)	(4.19e-06)	(3.60e-06)	(3.19e-06)
Age – 45	0.0124***	0.00294***	0.00124	0.00689***	-0.000445
	(0.00201)	(0.000787)	(0.00178)	(0.00164)	(0.000941)
$(\text{Age} - 45)^2$	-0.000135**	-6.38e-05***	4.99e-05	-0.000115***	3.91e-05
	(6.21e-05)	(2.34e-05)	(4.86e-05)	(4.16e-05)	(2.75e-05)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
Kleibergen-Paap Rank Wald	211.2	211.2	211.2	211.2	211.2
R-squared	0.051	0.011	0.032	0.064	0.015
Observations	7,742	7,742	7,742	7,742	7,742

Notes: This is the detail version of Table 10. It shows the estimated coefficients of other variables of the specifications presented in Table 10. Notes of Table 10 apply. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. The Effect of PM2.5 on Having Each Disease (2SLS)
The Second Stage Estimatoin: Non-farmers' Sample

Dependent variables	Hypertension	Diabetes	Lung Diseases	Heart Diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	0.00146	0.000795	0.000349	0.00325*	-0.000521
	(0.00272)	(0.00123)	(0.00151)	(0.00187)	(0.000715)
(Age – 45)	-0.000504	0.000171	7.80e-07	-0.000392**	0.000145
× (PM2.5 – meanPM2.5)	(0.000307)	(0.000164)	(0.000189)	(0.000184)	(9.37e-05)
$(Age – 45)^2$	1.98e-05**	-6.74e-06	2.38e-06	1.25e-05**	-5.10e-06
× (PM2.5 – meanPM2.5)	(9.30e-06)	(5.93e-06)	(6.54e-06)	(5.94e-06)	(3.20e-06)
Age – 45	0.0194***	0.00666***	0.000230	0.00851***	0.00184*
	(0.00331)	(0.00158)	(0.00208)	(0.00197)	(0.00106)
$(Age – 45)^2$	-0.000275***	-0.000115**	0.000109	-5.39e-05	-5.86e-07
	(9.62e-05)	(5.40e-05)	(6.82e-05)	(6.51e-05)	(3.58e-05)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
Kleibergen-Paap Rank Wald	137.8	137.8	137.8	137.8	137.8
R-squared	0.076	0.028	0.037	0.094	0.022
Observations	3,162	3,162	3,162	3,162	3,162

Notes: This is the detail version of Table 11. It shows the estimated coefficients of other variables of the specifications presented in Table 11. Notes of Table 11 apply. *** p<0.01, ** p<0.05, * p<0.1.

Table A4. The Effect of PM2.5 on Retirement : Robustness Checks
Using Province Level PM2.5 for Analysis : the Second Stage Results of 2SLS

Dependent Variables	Retirement Dummy			
Variables	(1)	(2)	(3)	(4)
Panel A. Farmers' Sample				
PM2.5 – meanPM2.5	0.00132	-0.00116	-0.000409	-0.00100
	(0.00204)	(0.00272)	(0.00368)	(0.00415)
(Age – 45)	-0.000151	-0.000131	-0.000133	-9.98e-05
× (PM2.5 – meanPM2.5)	(0.000318)	(0.000299)	(0.000302)	(0.000300)
$(Age – 45)^2$	-1.14e-07	-9.08e-07	-9.84e-07	-1.18e-06
× (PM2.5 – meanPM2.5)	(6.56e-06)	(6.49e-06)	(6.45e-06)	(6.31e-06)
Age – 45	-0.00365	-0.00344	-0.00345	-0.00348
	(0.00224)	(0.00221)	(0.00220)	(0.00222)
$(Age – 45)^2$	0.000598***	0.000602***	0.000606***	0.000605***
	(6.47e-05)	(6.54e-05)	(6.37e-05)	(6.38e-05)
Kleibergen-Paap Rank Wald	50.40	54.79	54.31	72.32
R-squared	0.192	0.210	0.212	0.219
Observations	7,742	7,742	7,742	7,742
Panel B. Non-farmers' Sample				
PM2.5 – meanPM2.5	0.00635**	0.00459	0.00834**	0.00783*
	(0.00250)	(0.00287)	(0.00350)	(0.00462)
(Age – 45)	-0.000862**	-0.000808*	-0.000818*	-0.000858**
× (PM2.5 – meanPM2.5)	(0.000423)	(0.000442)	(0.000446)	(0.000428)
$(Age – 45)^2$	2.17e-05*	1.97e-05	1.98e-05	2.12e-05*
× (PM2.5 – meanPM2.5)	(1.25e-05)	(1.31e-05)	(1.31e-05)	(1.26e-05)
Age – 45	0.0327***	0.0325***	0.0321***	0.0321***
	(0.00324)	(0.00346)	(0.00329)	(0.00315)
$(Age – 45)^2$	-0.000121	-0.000127	-0.000120	-0.000124
	(9.61e-05)	(0.000103)	(9.74e-05)	(9.23e-05)
Kleibergen-Paap Rank Wald	45.97	50.50	29.24	22.21
R-squared	0.351	0.358	0.366	0.369
Observations	3,162	3,162	3,162	3,162
Control Variables				
Individual Variables	✓	✓	✓	✓
Region Fixed Effect		✓	✓	✓
Province Variables			✓	✓
City Variables				✓

Notes:This is the detail version of Table 13. It shows the estimated coefficients of other variables of the specifications presented in Table 13. Notes of Table 13 apply. *** p<0.01, ** p<0.05, * p<0.1.

Table A5. The Effect of PM2.5 on Having Each Disease
Using Province Level PM2.5 for Analysis : the Second Stage Results of 2SLS
Farmers' Sample

Dependent variables	Hypertension	Diabetes	Lung Diseases	Heart Diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	-0.000704	-2.45e-05	0.00214	-0.000901	0.00170***
	(0.00217)	(0.000889)	(0.00159)	(0.00162)	(0.000661)
(Age – 45)	0.000274	8.67e-05*	-5.37e-05	7.02e-05	1.43e-05
× (PM2.5 – meanPM2.5)	(0.000296)	(4.53e-05)	(0.000164)	(0.000181)	(9.86e-05)
$(Age – 45)^2$	-9.88e-06	-2.66e-06*	3.35e-06	1.45e-06	1.86e-06
× (PM2.5 – meanPM2.5)	(9.73e-06)	(1.53e-06)	(3.90e-06)	(4.59e-06)	(3.09e-06)
Age – 45	0.0115***	0.00270***	0.00144	0.00618***	-0.000645
	(0.00214)	(0.000568)	(0.00157)	(0.00171)	(0.000900)
$(Age – 45)^2$	-0.000111*	-5.67e-05***	5.11e-05	-9.31e-05**	4.66e-05*
	(6.53e-05)	(1.78e-05)	(3.79e-05)	(4.72e-05)	(2.64e-05)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
Kleibergen-Paap Rank Wald	72.32	72.32	72.32	72.32	72.32
R-squared	0.051	0.011	0.030	0.062	0.014
Observations	7,742	7,742	7,742	7,742	7,742

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each province. PM2.5 is the average PM2.5 concentration in the past ten years in each province. meanPM2.5 is the national average of PM2.5 in the past ten years. Column (1)-(5) show different estimation results of different regression which use each type of disease dummy as a dependent variable. *** p<0.01, ** p<0.05, * p<0.1.

Table A6. The Effect of PM2.5 on Having Each Disease
Using Province Level PM2.5 for Analysis : the Second Stage Results of 2SLS
Non-farmers' Sample

Dependent variables	Hypertension	Diabetes	Lung Diseases	Heart Diseases	Asthma
Variables	(1)	(2)	(3)	(4)	(5)
PM2.5 – meanPM2.5	-0.000739	0.00308	-0.00172	0.00209	0.000263
	(0.00451)	(0.00219)	(0.00197)	(0.00273)	(0.00121)
(Age – 45)	-0.000244	0.000362**	0.000214	-0.000227	0.000128
× (PM2.5 – meanPM2.5)	(0.000171)	(0.000177)	(0.000166)	(0.000197)	(0.000139)
$(Age - 45)^2$	1.36e-05***	-1.29e-05*	-5.17e-06	9.16e-06	-3.96e-06
× (PM2.5 – meanPM2.5)	(4.83e-06)	(6.75e-06)	(7.04e-06)	(6.89e-06)	(4.87e-06)
Age – 45	0.0181***	0.00643***	-0.000197	0.00759***	0.00204*
	(0.00233)	(0.00154)	(0.00189)	(0.00211)	(0.00119)
$(Age - 45)^2$	-0.000233***	-0.000110**	0.000128*	-2.78e-05	-8.95e-06
	(6.17e-05)	(5.00e-05)	(6.68e-05)	(6.91e-05)	(3.87e-05)
Control Variables					
Individual Variables	✓	✓	✓	✓	✓
Region Fixed Effect	✓	✓	✓	✓	✓
Province Variables	✓	✓	✓	✓	✓
City Variables	✓	✓	✓	✓	✓
Kleibergen-Paap Rank Wald	22.21	22.21	22.21	22.21	22.21
R-squared	0.075	0.029	0.037	0.093	0.022
Observations	3,162	3,162	3,162	3,162	3,162

Notes: Clustering robust standard errors in parentheses assuming that the error terms are correlated within each province. PM2.5 is the average PM2.5 concentration in the past ten years in each province. meanPM2.5 is the national average of PM2.5 in the past ten years. Column (1)-(5) show different estimation results of different regression which use each type of disease dummy as a dependent variable. *** p<0.01, ** p<0.05, * p<0.1.

Table A7. The Effect of PM2.5 on Retirement (2SLS): the Second Stage Results Dropping Respondents who Moved to the Current Place after 2004

Dependent variable	Retirement Dummy			
Variables	(1)	(2)	(3)	(4)
Panel A. Farmers' Sample				
PM2.5 – meanPM2.5	0.00216**	0.00185	0.00217	-3.47e-05
	(0.00104)	(0.00126)	(0.00153)	(0.00219)
(Age – 45)	-0.000356**	-0.000319*	-0.000316*	-0.000294*
× (PM2.5 – meanPM2.5)	(0.000170)	(0.000172)	(0.000172)	(0.000173)
$(Age - 45)^2$	2.77e-06	1.82e-06	1.70e-06	1.87e-06
× (PM2.5 – meanPM2.5)	(5.51e-06)	(5.30e-06)	(5.28e-06)	(5.19e-06)
Age – 45	-0.00265	-0.00287	-0.00289	-0.00283
	(0.00214)	(0.00215)	(0.00216)	(0.00220)
$(Age - 45)^2$	0.000597***	0.000606***	0.000611***	0.000606***
	(6.51e-05)	(6.48e-05)	(6.46e-05)	(6.50e-05)
Kleibergen-Paap Rank Wald	154.8	212.2	208.6	214.1
R-squared	0.199	0.213	0.215	0.221
Observations	7,666	7,666	7,666	7,666
Panel B. Non-farmers' Sample				
PM2.5 – meanPM2.5	0.00496**	0.00464**	0.00650***	0.00656**
	(0.00207)	(0.00217)	(0.00240)	(0.00283)
(Age – 45)	-0.000717**	-0.000685**	-0.000696**	-0.000727**
× (PM2.5 – meanPM2.5)	(0.000301)	(0.000304)	(0.000301)	(0.000300)
$(Age - 45)^2$	1.75e-05*	1.60e-05*	1.67e-05*	1.78e-05*
× (PM2.5 – meanPM2.5)	(9.19e-06)	(9.31e-06)	(9.17e-06)	(9.13e-06)
Age – 45	0.0337***	0.0335***	0.0333***	0.0332***
	(0.00399)	(0.00402)	(0.00392)	(0.00389)
$(Age - 45)^2$	-0.000147	-0.000148	-0.000148	-0.000152
	(0.000109)	(0.000112)	(0.000108)	(0.000107)
Kleibergen-Paap Rank Wald	96.55	100.2	149.1	138.7
R-squared	0.351	0.358	0.365	0.368
Observations	3,052	3,052	3,052	3,052
Control Variables				
Individual variables	✓	✓	✓	✓
Region fixed effect		✓	✓	✓
Province variables			✓	✓
City variables				✓

Notes: The sample is restricted to respondedents who lived even before 2004 in the current place. Notes of Table 9 apply. *** p<0.01, ** p<0.05, * p<0.1.