Air Pollution and Migration: Evidence From China’s Dynamic Monitoring Survey Data

Yantuan Yu, Guangdong University of Foreign Studies, China
Weidong Zhao, Shanghai Lixin University of Accounting and Finance, China
Xiuzhen Li, Shanghai Lixin University of Accounting and Finance, China
Yun Zhang, Shanghai Lixin University of Accounting and Finance, China*

ABSTRACT

This paper investigates the impact of air pollution on China’s internal migration using the migrant population dynamic monitoring survey and air quality index (AQI) data in 2014. Binary response models suggest that, on average, the probability of the willingness to leave will grow by approximately 0.1 if the average value of AQI is increased by 100 points. The migration effect of severe air pollution still holds using two-stage least squares estimation. In particular, compared to inter-province and intra-city within the same province, interviewees are more willing to leave intra-county with the same city. Individuals from the cities in central and western regions, the listed key environmental protection (KEP) cities, and cities not listed as resource-based (RB) cities have stronger willingness to leave than those from the cities in eastern region, non-KEP cities, and RB cities. The results are robust to alternative measures of air pollution and model specification.

KEYWORDS

Air Pollution, Big Survey Data, Binary Response Models, Internal Migration, IV Estimation

1. INTRODUCTION

A satisfy livable environment is an important condition for the city to continuously attract foreign talents and ensure sustainable prosperity. Based on the questionnaire survey data, for example, subjective acceptance of air quality has a significant positive impact on the willingness of graduate students to stay in Beijing (Hao et al., 2020). The ecological environment is the basis for human survival and economic development. However, economic growth is excessively dependent on resource consumption with high pollution and high emission. Besides, the coal-based energy consumption structure and more and more motor vehicle possession bring about continuous deterioration of the ecological environment (Chen et al., 2013), especially the air pollution has become a prominent dilemma in China’s ecological environmental governance. According to the “China Environmental Status Bulletin” in 2015, 265 cities in 338 prefecture-level cities and above had excessive ambient air quality, accounting for 78.4%, and the average over-the-counter ratio was 23.3%. The PM2.5 concentration (fine particulate matter with diameters equal or smaller than 2.5μm) is the most
prominent air pollutant in Chinese cities and it is also the main component of Air Quality Index (AQI). According to the international environmental protection organization, in 2015, the PM2.5 annual average concentration of China’s 366 cities is 50.2 μg/m³. However, 80% of them has not yet reached the China’s ambient air quality standards.

With the development of urbanization and industrialization in China, the ecological environment problem becomes more and more serious, especially for air pollution (Rafiq et al., 2016). Not only severe environmental pollution brings great troubles to people’s ecology, production and living space, but also causes the decline of residents’ health and overall welfare levels and the increases the incidence rate of related diseases. Besides respiratory diseases and cancer, it has also greatly affected the prevalence rate of cardiovascular and cerebrovascular diseases (Xu et al., 2013), and has become one of the major threats to people’s health in China (Yang et al., 2013). Air pollution not only brings harm to our body and mind, but also brings inconvenience to our life. The smog in January 2013 affected 1.4 million square kilometers of our country and brought adverse effects to more than 800 million people (Xu et al., 2013). Meanwhile, environmental pollution has also increased the expenditure burdens, exerting adverse effects on economic and social development. During the 12th Five-Year Plan period, the Chinese government invested as much as 3.4 trillion yuan in environmental protection (Xu et al., 2013), accounting for 1.4% of Gross Domestic Product.

Previous studies have found that air pollution can lead to increased depression and suicidal behavior (Lim et al., 2012). With the worsening of urban environmental pollution, some vulnerable people who are in poor health will choose to leave heavily polluted cities for fear of health damage. Some studies confirm that air pollution is one of the reasons for the changes in the inflow and outflow of Chinese immigrants by using the national county population census and 1% sampling data. Specifically, the independent changes in China’s air pollution from 1996 to 2010 reduced the population inflow by 50%. Air pollution causes a net outflow of 5% of the population for a specific county (Chen et al., 2017). Integrating natural environment into the study of population migration has become an important direction of migration research. Migration will affect the exposure level of migrant population to PM2.5. Shen et al. (2017) found that the total pollution exposure level of the population decreased by 52μg/m³ on average from 2000 to 2010 by using China’s demographic data and household fuel usage, of which 60% could be attributed to internal migration and flow. In 2010, the national average exposure concentration of PM2.5 decreased by 3.9μg/m³, equivalent to a reduction of 36,000 premature deaths per year. In the 1960s and 1970s, severe smog pollution broke out in Los Angeles area of California. After the 1970s, the state government decided to carry out smog removal operations. Consequently, it was found that the population in relatively poor areas, increased significantly with the reduction of smog. On the one hand, the settlement of smog problem has significantly increased people’s willingness to live and reduced the outflow rate of population. On the other hand, as the demand for labor increases, the number of immigrants increases, and the region changes from the place where the population flows out to the place where the population flows in (Kahn, 2000). In other words, people can avoid the harsh living environment through migration. Besides, daily air pollution levels have a significant effect on the decision to purchase or cancel health insurance (Chang et al., 2018).

With the continuous improvement of living standards, the people have more and more urgent demands for fresh air, clean drinking water, healthy food and a beautiful environment. When the residence faces prominent environmental pollution, the people’s choice of residence may have the following three situations (Reuven, 2007). First, they will continue to stay in the original city of residence, accept the welfare losses caused by environmental pollution, and will not take any measures. Continuing to settle in the city has many benefits, such as more job opportunities, better social and public services, and even psychological satisfaction brought about by living in a big city. Residents believe that the losses caused by environmental pollution can be compensated by the above benefits. Second, we will continue to stay in the original city of residence, but we will reduce the possible losses caused by environmental pollution through various means, such as reflecting our own needs through
various means, exerting pressure on the government to evacuate polluting enterprises. In addition, residents may also reduce the impact of environmental pollution on themselves by purchasing air purifiers, masks and other equipment spontaneously. Third, take the initiative to move. Due to the deterioration of the city’s environmental conditions, the degree of environmental risks faced by people has changed, as well as the scale and quality of ecological services that can be provided locally, thus affecting the welfare of residents. A good urban ecosystem can provide reliable ecological products to residents, enrich the connotation of urban culture, form spontaneous regulations on environmental damage, and comprehensively improve the well-being of residents (Black et al., 2011; Fisher et al., 2009), when the urban ecosystem is challenged, the above benefits will gradually disappear and the welfare level of residents will tend to decline. Settlement and migration of population are two aspects of the same problem. If residents’ environmental demands are not satisfied, then when facing increasingly prominent environmental problems, some people will seek to take the initiative to move out to avoid adverse effects.

Labor plays an extremely important role in the sustainable development of the city, and the migration of population actually represents the migration of labor force. For instance, higher PM2.5 concentration drives college graduates away from their college city (Lai et al., 2021). The movement of population between cities will have an important impact on the long-term development of each city. Generally speaking, the outward flow of population will reduce the level of human capital in the places where the population is exported, which is relatively unfavorable to the development of the places where the population is exported, and the continuous output of population will make it difficult for cities to exert the economies of scale brought about by public services and infrastructure construction. At the same time, the migration of labor force between regions has caused the gap between regions to narrow or expand, bringing about both advantages and disadvantages in the economic development of each region, which is an important factor affecting the urbanization process and economic and social development of each city.

As the environmental pollution, especially the air pollution, in various cities of our country has increasingly become an influential factor that cannot be ignored in the decision-making of residents’ settlement and migration, it is of great practical significance to discuss to what extent the environmental pollution in cities will have an impact on the decision-making of residents’ settlement and to analyze under what circumstances it will have an impact. Judging from the existing literature, the issue of population migration caused by environmental pollution has received less attention from academia, especially the impact of urban environmental pollution on population migration, which has hardly been involved in either domestic or foreign historical studies. For this reason, this study uses the matching data of China’s urban air quality and the Migrant Population Dynamic Monitoring Survey Data to discuss the impact of environmental pollution on population migration (settlement).

Our study contributes to the literature in two ways. On the one hand, based on the structure framework of “push-pull” theory (Lee, 1966), we develop a theoretical model to investigate how migration effects occur. On the other hand, we examine the heterogeneity of the effects of air pollution on migration intention in areas with different economic development levels and migration ranges (inter-province, intra-city within the same province and intra-county within the same city). It is found that residents in economically underdeveloped areas are more likely to have migration behavior caused by non-economic factors, which enriches the research on the influence of air pollution on migration intention had provides references for cities to cope with air pollution sustainably.

The paper proceeds as follows. Section 2 summarizes related literature. Section 3 outlines the theoretical framework. Section 4 describes the econometric strategy and data. Section 5 illustrates and discusses the results. The final section concludes this study.
2. RELATED LITERATURE

An increasing body of literature has focused on environmental pollution and migration. So far, this literature has consisted of two main strands: (1) the strand covering the definitions of environmental refugees and environmental migration, along with classifying the environmental pollution; (2) empirically, the strand analyzing how environmental pollution especially air pollution affects migration directly and indirectly.

The literature of the first strand, i.e., the concepts of environmental refugees (El-Hinnawi, 1985) and environmental migration (Myers, 1997; Bates, 2002). The impacts of environmental pollution on migration are usually involving these two definitions. According to the International Organization for Migration (IOM, 2009), environmental migrants refer to those who voluntarily or are forced to leave their original place of residence temporarily or permanently due to sudden or gradual environmental deterioration (Piguet, 2010), including both voluntary migrants and forced migrants, including both temporary migrants and permanent migrants, including both domestic migrants and international migrants. Immigrant studies triggered by early environmental problems mainly appeared in the fields of geography and sociology (Millock, 2015). The focus is on human migration and social evolution caused by changes in geology, climate or natural disasters, etc. The linkage between environmental pollution and emigration are also observed in recent years (Qin and Zhu, 2018; Liu and Yu, 2021).

Many types of environmental problems may lead to migration, and different environmental problems may have different impacts on population migration. Therefore, it is necessary to classify environmental problems according to certain principles. Generally, environmental problems can be divided into two categories according to the length of time they occur. One is sudden disasters, such as earthquakes, floods, hurricanes, extreme weather, epidemics (Mbaye and Zimmermann, 2016). The other is gradual environmental degradation, such as soil degradation, sharp decrease in vegetation, global warming (Gray and Mueller, 2012; Beine and Parsons, 2015). The difference between the two is that the occurrence of the first kind of environmental problems is sudden. Although advanced monitoring techniques can be used to detect signs in some developed countries, the vast majority of affected residents have no time to respond when facing such environmental problems. Only after the disaster, depending on the severity of the disaster, can they choose to restore to the original state or move permanently to other places. The occurrence of the second kind of environmental problems is chronic. Local residents can have some experience of these problems in a long time and will take corresponding measures to cope with the deteriorating situation. There is a big difference in the impact of these two environmental problems on immigrants. According to the role of human beings in environmental problems, environmental problems can be divided into three categories. One category is basically unrelated to human activities, such as large earthquakes, volcanic eruptions, tsunamis, etc. The other category is related to human activities, but not entirely caused by human activities, such as ozonosphere hole, global warming, etc. The third category is completely caused by human activities, and all types of environmental pollution belong to this category of environmental problems. According to the relationship between environmental changes and individual or family migration decisions, it can be divided into endogenous environmental changes and exogenous environmental changes. The former refers to environmental problems closely related to economic and social factors such as forest degradation, soil erosion and water quality decline, while the latter refers to environmental problems such as volcanic eruptions, rainfall changes and extreme climate that do not depend on human activities at least in recent decades (Millock, 2015).

In the face of sudden natural disasters such as floods, hurricanes and earthquakes, immigration is a self-protection mechanism (Bousitan et al., 2012), which can effectively avoid risks and reduce losses. Carvajal and Medalho Pereira (2009) studied the problem of migration in Nicaragua after the hurricane in 1998 and found that since houses were hard to bear the impact of heavy rain after the hurricane, the affected families had to carefully consider whether to decide to relocate, but the disaster itself did not guarantee the relocation of families in the next few years. However, only
when the family has enough wealth can the family members complete the actual migration. For the relatively poor families in rural areas, they are more affected by the rainstorm, but the probability of migration is lower, which is exactly the opposite in urban families. In fact, sudden environmental problems, such as earthquakes and volcanic eruptions, will also destroy residents’ property and thus reduce the ability of families to migrate outward, which is also called “mobility restriction”. At the same time, these sudden environmental problems will also destroy the transportation infrastructure network such as roads and railways, making it difficult to migrate in the short term. Halliday (2006) found that the earthquake had a negative impact on immigrants, for both rich and poor families. The possible reason is that multiple constraints affect migration exist at the same time. In addition, since rebuilding homes after the earthquake requires a large amount of labor force, the disaster area has a large inflow of labor force in a certain period of time (Alexeev et al., 2011).

At the same time, many studies have found that some sudden environmental problems have not brought about immigration. For example, Paul (2005) used survey data from eight villages hit by tornadoes in Bangladesh to find that due to effective government and private assistance, no one moved to other areas. Boussan and Kahn (2012) also found that the flood control plan implemented in the United States in the decade of the 20th century explained that the flood-hit states in the following decades brought a net inflow of population, while the tornado-hit states brought a net outflow of population.

Perch-Nielsen et al. (2008) found that flooding is difficult to cause population migration, even if there is, it is only temporary. Sudden environmental problems are not easy to cause permanent migration. Bohra-Mishra et al. (2014) used the sum of Indonesian household panel data at the provincial level, it is found that except for landslides, various types of environmental problems including earthquakes, floods and volcanic eruptions have no significant impact on permanent migration.

The new labor migration economics thinks that the family is the basic unit to decide the labor force decision, and migration is regarded as a strategy of risk diversification. Gradual environmental problems mainly include slow changes in temperature and rainfall, drought and soil salinization. Compared with sudden environmental problems, the main difference in the impact of gradual environmental problems on population migration is that the latter has a longer time to choose coping strategies, compare the benefits of migration or continued settlement, and then make the final decision. Gutmann et al. (2005) found that changes in temperature and rainfall in the 1930s were an important driving force for population migration in counties and districts on the Great Plains of the United States. Gray (2009) found that the increase in average annual rainfall did not cause local population migration, but had a significant negative impact on internal cross-regional migration and international population migration. At the same time, Gray (2009) also believed that wealth played an important role in the long-distance migration of residents. In addition, some economic studies usually regard such gradual environmental changes as instrumental variable to identify impacts, and indirectly discover the impact of such environmental problems on population migration in the process. For example, Beegle et al. (2011) found that rainfall has a significant impact on population migration when taking rainfall as instrumental variable.

Gray (2011) adopted the follow-up survey data from Kenya and Uganda, it is found that soil quality is significantly negatively related to migration in Kenya, especially for temporary labor flows, but the opposite is true in Uganda. It confirms that adverse environmental conditions will increase migration. Di Falco et al. (2011) employed household survey data in the Nile Basin of Ethiopia and found that less than 5% of households choose immigration as a coping strategy in the face of climate change. Koubi et al. (2016) suggested that sudden environmental events, such as storms or floods, usually increase the possibility of environmental migration, but also pointed out that long-term and gradual environmental changes, such as drought and soil salinization, will give local residents enough time to cope with, but will reduce the possibility of migration. Still some studies have simulated the impact of climate change on future migrants at the national level, such as Hassani-Mahmoei and
Parris (2012). Moreover, residents close to hazardous waste landfills will tend to choose migration as long as the per capita income exceeds the threshold (Gawande et al., 2000)

3. THEORETICAL FRAMEWORK

The cross-regional flow of population needs to consider the simultaneous action of the pushing force in the region where the actor is located and the pulling force outside the region. The “push-pull” model is a classic model in the immigration decision theory (Todaro, 1969; Borjas, 1989), this model assumes that the migration of population is the result of the combined action of the pushing force of the destination and the pulling force of the destination, and the decision of the migration is made by the actor after comparing the potential benefits and costs of the migration. According to this idea, and drawing on the research of Reuven and Moore (2009), it is assumed that potential migrants will make a decision on whether to migrate by comparing the net income before and after migration.

If the actor continues to live in his current place, he may have different expected income according to his personal characteristics, family characteristics and the city where he is located. Assuming that the expected income equation is in the form of Mincer equation, the expected revenue \((ER)\) can be expressed as follows:

\[
\ln ER_f = E_j + w_f X + Y_f + \mu_f, \mu_f - N \left(0, \sigma_f^2\right)
\]

(1)

where \(ER_f\) is the expected revenue of the specific individual in the focal place; \(E_j\) is the focal environmental conditions of the residence; \(X\) denotes personal and family characteristics; \(Y_f\) represents the focal public services and other urban characteristics which not affected by personal influence; \(w_f\) is the coefficient of personal and family characteristics; \(\mu_f\) is the random variable which captures the personal adaptability, popularity, luck and other uncontrollable factor’s influence on internal migration.

Similarly, if the individual chooses to move to another city (assuming that the individual will first make a comparison among all the cities and choose the city that is most beneficial to him according to the principle of maximizing the \(ER\)), the \(ER\) specified as:

\[
\ln ER_{-,j} = E_{-,j} + w_{-,j} X + Y_{-,j} + \mu_{-,j}, \mu_{-,j} - N \left(0, \sigma_{-,j}^2\right)
\]

(2)

where \(ER_{-,j}\) denotes the \(ER\) of the neighboring cities of the focal city. Other symbol interpretation is similar to Model (1).

Assuming the correlation coefficient between \(\mu_f\) and \(\mu_{-,j}\) is \(\rho\), which ranges from -1 to 1. If the individual can bring positive \(ER\) from unobservable factors such as adaptability, popularity, and luck in the current place of residence, then the same is true in other cities after migration, while \(\rho\) is positive and close to 1. However, the unobservable factors have the characteristic of city dependence if \(\rho\) is negative. That is to say, the individual can bring positive benefits through these factors and once he decides to migrate, these factors cannot play a role in neighboring cities. Generally speaking, random variables representing unobservable factors can play a similar role in different cities, i.e., \(\rho\) is generally greater than zero.

The individual will compare the \(ER\) between ex ante and ex post internal migrations. In addition, due to the existence of transportation costs and other potential opportunity costs, the actual relocation needs to compare the \(ER\) of the ex post migration and the \(ER\) of the ex ante migration plus costs. The \(ER\) gap \((ERG)\) is defined as:
ERG = ln\left(\frac{ER_j}{ER_j + C}\right) = (E_{-j} - E_j) + (w_{-j} - w_j)X + (\mu_{-j} - \mu_j) - \pi \tag{3}

where \pi = ln\left(\frac{ER_j + C}{ER_j}\right) \equiv \frac{C}{ER_j}, if \frac{C}{ER_j} is as small as possible the relationship between \equiv is satisfied.

The condition for the individual to decide to move from his current place of residence to other cities is that the ER gap is positive, i.e., ERG > 0. Judging from the components on the right side of Model (3), the realization of this condition is easier to hold when the following constraints are satisfied:

1. $E_{-j} - E_j > 0$, that is, the migrant’s environmental benefits after moving to settle in the new city are greater than those in the current city, or the individuals are preferred to move to cleaner cities. Intuitively, it is easy to understand that the city has a good ecological environment to make people happy physically and mentally, and the residents have a higher happiness index, which can better attract the migration population to settle down.

2. $w_{-j} - w_j > 0$. $w_j$ and $w_j$ are the coefficients of personal and family characteristics after migration and when they continue to live in their current place respectively. Individual factors or even all factors such as individual’s age, education level, work experience and family income may have different returns between the two places. In order for the migration behavior of the actor to occur, these factors need to be weighted to have higher economic returns in the migrated city. For example, when the number of years of education is fixed, the migration population will usually choose cities with higher educational returns to settle down. Due to many factors of these personal and family characteristics, some factors will bring positive income returns, while others will be opposite, so the weighted returns of these factors need to be considered.

3. Since $\mu$ is used to measure other uncontrollable factor such as personal adaptability, popularity, luck, etc. For a specific individual, if $\mu_j > 0$, then requires $\rho > 0$ and tends to 1, thus $\left(\mu_{-j} - \mu_j\right) \to 0^+$, that is to say, if the individual’s income from other uncontrollable factor such as personal adaptability, popularity, luck, etc. in his current place of residence is positive, the migration is only occurs if the income is also positive in the migration destination. On the other hand, if $\mu_j < 0$, then if $\rho < 0$ is satisfied, migration is more likely to occur.

4. $\pi \to 0^+$, because $\pi$ is related to costs, to meet the above conditions, the transportation costs and other implied opportunity costs of the actor’s relocation are required to be as low as possible.

Given that, if the above four constraints are met or some of them are met and sufficient to lead to ERG > 0, the individual will decide to move from his current place of residence to another city.

4. ECONOMETRIC STRATEGY AND DATA

4.1 Econometric Models Specification

A generalized linear model for binary response data has the form

$$\Pr(y = 1 \mid x) = g^{-1}\left(x'\beta\right)$$
where \( y \) is the 0/1 response variable, which denotes internal migration (equals to 1) or not (equals to 0), \( x \) is the \( n \)-vector of predictor variables containing the interested variable air pollution, environmental-, social-, economical-, and political factors, \( \beta \) is the vector of regression coefficients, and \( g \) is the link function.

In the existing literature, the most common choices for the link function are Logit, Probit and complementary log-log, which specified as follows:

\[
g(p) = \log \left( \frac{p}{1-p} \right) \quad (5)
\]

\[
g^{-1}(\eta) = \Phi(\eta) \quad (6)
\]

where \( \Phi \) is the cumulative distribution function for the standard normal distribution.

\[
g(p) = \log(-\log(1-p)) \quad (7)
\]

All three of these are strictly increasing, continuous functions with \( g(0) = -\infty \) and \( g(1) = +\infty \).

We will discuss when to use each of these link functions. The Probit link function is appropriate when it makes sense to think of the response \( y \) as obtained by thresholding a normally distributed latent variable \( z \):

\[
z = x'\beta^* + \varepsilon \quad (8)
\]

where \( \varepsilon \sim N(0, \sigma^2) \) and \( y = 1 \) if \( z \geq 0 \), \( y = 0 \) otherwise.

Defining \( \beta = \frac{\beta^*}{\sigma} \), these yields

\[
\Pr(y = 1 \mid x) = \Pr(x'\beta^* + \varepsilon \geq 0) = \Phi(x'\beta^*) \quad (9)
\]

Logit model is the default link function to use when we have no specific reason to choose one of the others. There is a specific technical sense in which use of Logit corresponds to minimal assumptions about the relationship between \( y \) and \( x \). Suppose that we describe the joint distribution for \( x \) and \( y \) by giving

- the marginal distribution for \( x \), and
- the expected value of \( x_i y \) for each predictor variable \( x_i \).

Then the maximum-entropy (most spread-out, diffuse, least concentrated) joint distribution for \( x \) and \( y \) satisfying the above description has a probability density function of form

\[
p(x, y) = \frac{1}{Z} f(x) \exp \left( \sum_{i=1}^{n} \beta_i x_i y \right) \quad (10)
\]
for function $f$, coefficient vector $\beta$ and normalizing constant $Z$. The conditional distribution for $y$ is then

$$p(y | x) = \frac{p(x, y)}{p(x, 0) + p(x, 1)} = \frac{\exp\left((x'\beta)y\right)}{1 + \exp(x'\beta)}$$

(11)

and so

$$\Pr(y = 1 | x) = \frac{\exp(x'\beta)}{1 + \exp(x'\beta)} = \log^{-1}(x'\beta)$$

(12)

The complementary log-log link function arises when

$$y = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{if } z = 0 \end{cases}$$

(13)

where $z$ is a count having a Poisson distribution:

$$z \sim \text{Poisson}(\lambda), \lambda = \exp(x'\beta)$$

(14)

To see this, let

$$p = \Pr(z > 0 | x)$$

(15)

Then

$$p = 1 - \text{Poisson}(0 | \lambda) = 1 - \exp(-\lambda) = 1 - \exp(-\exp(x'\beta))$$

(16)

and so

$$cloglog(p) = \log\left(-\log(1 - p)\right) = \log\left(-\log\left(\exp\left(-\exp(x'\beta)\right)\right)\right) = x'\beta$$

(17)

In this paper, to obtain the robustness results, we adopt three link functions discussed above to carry out empirical analysis except for the ordinary least squares (OLS).

4.2 Data and Variables

4.2.1 Data Descriptions

This study is carried out at the individual level. Comparing and analyzing the large-scale micro survey data in China, the dynamic monitoring survey data of migrant population provides us a good opportunity to observe the migration decision. The survey data in 2014 is organized and carried out
by the National Health and Family Planning Commission of China (NHFPC). The sample points were distributed in 1459 county-level unites in 31 provinces, autonomous regions and municipalities and Xinjiang Production and Construction Corps, involving 3776 subdistricts and towns (townships) and 8993 neighborhoods (villagers’ committees). Interviewees were the migrant population who had lived in the inflow places for more than a month, did not have household registration in that city (district or county), and were at the age of 15-59 by May 2014. Those whose current residence if different from the place of their household registration in the same city were not included. Sample sizes of the provinces were divided into the eighth grades of 14000 persons, 12000 persons, 1000 persons, 8000 persons, 7000 persons, 6000 persons, 5000 persons and 4000 persons. According to the survey plan, 201000 persons were to be interviewed, but 200937 persons were actually interviewed, involving 667122 family members, of which, 575288 family members were in the inflow places. Contents under monitoring include basic information of migrant population and family; employment and housing; basic public health and medical service; conditions about marriage and childbearing and family planning services; basic condition of community population; community management and services and so on. Results of the survey have well represented not only the whole county and the provinces, but also urban groups and key cities. Moreover, to improve air quality, China issued a new version of Ambient Air Quality Standards in 2012, and the first survey data issued is in 2014, so we choose the survey data in 2014 as the study case.

The main reasons for selecting this dataset are as follows. On the one hand, the dataset provides rich individual and family characteristics of residents in various regions of China, which can comprehensively measure the above-mentioned social, economic and other influencing factors. On the other hand, compared with native residents, migrant population has fewer fixed assets, simpler social network and lower opportunity cost to make migration decisions. Therefore, migrant population is more likely to migrate. Finally, selecting migrant population as the research object can avoid the rare event bias.

4.2.2 Variable Definitions

Dependent variable. The dependent variable of this article is the long-term residence willingness, i.e., “whether to plan to live here for more than five years”. There are three options for this question. About 56% of the interviewees said that they would continue to live for more than five years (willingness to stay, wts), only 13.5% did not plan to live for a long time (willingness to leave, wtl), and about one third said they did not think well which deletes directly for the sake of facilitating the study and only analyze those with clear residence intention. The dependent variable or binary responses assigned to 1 for wtl and 0 for not wtl (wts). Due to the partial absence of independent variables, 105878 samples were finally included in the analysis.

Interested variables. This paper utilizes the year average level of AQI daily data (maqi) published by the Ministry of Ecology and Environment of the People’s Republic of China to reflect the air pollution situation in the city where the individual lives. The AQI focuses on health effects one may experience within a few hours or days after breathing polluted air. Environment Protection Agency (EPA) calculates the AQI for five major air pollutants regulated by the Clean Air Act: ground-level ozone, particle pollution (also known as particulate matter), carbon monoxide (CO2), sulfur dioxide (SO2), and nitrogen dioxide (NO2). For each of these pollutants, EPA has established national air quality standards to protect public health. Ground-level ozone and airborne particles are the two pollutants that pose the greatest threat to human health. As for the robustness checks, the total days with AQI>200, PM2.5 and industrial SO2 emissions are employed to alternative measure air pollution intensity.

Control variables. In addition to air pollution, other family and personal variables that affect internal migration decisions include family monthly income (fincome), household nature (hnature), age (age), gender (gender), education level (education), marital status (mstatus), nationality (nation),
sample point type (spoint), the time from the last migration (leave), whether to buy a house or own a house (house), whether to have endowment insurance (uwd), house accumulation fund (haf), and new agricultural insurance (nai). In addition, the individual’s decision-making is also influenced by the city characteristics of the city, especially the public service and infrastructure characteristics. Such the house price of the focal city (price), logarithm value of the total population for each city in 2014; the population density (popd), proxied by the share of 10,000 persons in city area (km²), we control for population density because it may affect air pollution (Chen et al., 2020a); public services, the share of 10,000 books and doctors in the total population (10,000 persons); removal rate of industrial wastes including solid waste (rsw) and waste water (rww). We could not include other factors, such as technological changes (Chen et al., 2020b; Chen et al., 2020c), because of this factor may exert more impacts on pollution (or pollution intensity) than labor mobility. Summary statistics are reported in Table 1.

Table 1. Summary statistic

| Variables | Observations | Mean | Std. Dev. | Min. | Max. |
|-----------|--------------|------|-----------|------|------|
| wtl       | 105878       | 0.2006 | 0.4004   | 0.0000 | 1.0000 |
| maqi      | 100616       | 98.5708 | 23.9373 | 51.0000 | 226.0000 |
| fincome   | 93667        | 8.0384 | 0.5569   | 0.6931 | 14.2210 |
| hnature   | 105878       | 0.1415 | 0.3486   | 0.0000 | 1.0000 |
| age       | 105862       | 33.9380 | 9.2323  | 15.0000 | 59.0000 |
| gender    | 105862       | 0.5832 | 0.4930   | 0.0000 | 1.0000 |
| education | 105862       | 3.3613 | 0.9922   | 1.0000 | 7.0000 |
| mstatus   | 105861       | 1.8743 | 0.5296   | 1.0000 | 5.0000 |
| nation    | 105878       | 0.9308 | 0.2538   | 0.0000 | 1.0000 |
| spoint    | 105862       | 0.2859 | 0.4519   | 0.0000 | 1.0000 |
| leave     | 105862       | 9.7312 | 6.5192   | 1.0000 | 51.0000 |
| house     | 105878       | 0.2082 | 0.4061   | 0.0000 | 1.0000 |
| uwd       | 105878       | 0.2328 | 0.4226   | 0.0000 | 1.0000 |
| haf       | 105878       | 0.0757 | 0.2644   | 0.0000 | 1.0000 |
| nai       | 105878       | 0.5053 | 0.5000   | 0.0000 | 1.0000 |
| price     | 105877       | 0.8500 | 0.4700   | 0.2900 | 2.4400 |
| popd      | 105878       | 0.0516 | 0.0347   | 0.0006 | 0.2648 |
| books     | 104844       | 108.2780 | 125.5264 | 5.5700 | 920.0300 |
| doctors   | 104071       | 93.4673 | 13.5433 | 5.4900 | 157.9400 |
| rsw       | 104790       | 86.0352 | 18.7457 | 6.0000 | 100.0000 |
| rww       | 102949       | 86.8207 | 11.2479 | 22.3100 | 100.0000 |
5. RESULTS AND DISCUSSION

5.1 Baseline Results and Identification

To examine the impact of air pollution on internal migration, empirical analysis was carried out by using Probit, Logit, CLog-Log and OLS estimations. The baseline results are shown in Table 2. Empirical results show that the average value of AQI in the city is significantly and positively correlated with the internal migration willingness, indicating that the poorer the air quality in the city, the willingness to leave ($wtl$) of the focal residents are stronger. Specifically, a 100 point increase in the average value of AQI would lead to an increase in willingness to leave by nearly 10%, *ceteris paribus*. This conclusion still holds after including family, personal and city characteristics.

Among the variables of family and personal characteristics, the average monthly total income of the family in the local area ($fincome$) is negatively correlated with $wtl$. This result shows that if the family’s income in the local area is higher, the greater the stickiness to the individual in the city where the family is located, the lower the possibility of choosing migration, and in turn, the greater the possibility of planning to live permanently in the local area. The older the individual is ($age$), the lower the possibility of settling down and the greater the possibility of migration, which is related to the characteristics of the internal migration. The longer it is from the first movement ($leave$), the more likely it is to continue to settle in the local area. The possible reason is that the individual has been able to adapt well to the city where he is located, and the decision to move requires greater external costs. Non-agricultural registered permanent residence ($hnature$), gender (male), Han nationality ($nation$), sampling point are the village committees ($spoint$), house accumulation fund ($haf$) and new agricultural insurance ($nai$) that will improve $wtl$ since the coefficients of these variables are significantly positive. Owning purchased and self-built houses and pension insurance for urban workers will reduce residents’ $wtl$.

As for the characteristic variables of the city, the higher the house price in the city ($price$), the more likely the residents will choose to live in the local area. This result seems to be inconsistent with the expectation. However, it is necessary to consider that the influence of house price on $wtl$ may have three effects. First, the continuously rising house price will weaken the resident’s willingness to continue to settle in the focal area and form the house price driving effect. Second, considering that higher housing prices in cities usually indicate better infrastructure and public services, and also imply more opportunities for host cities to provide, thus attracting migration population to continue to settle in the local area, this characteristic of cities is a welfare magnet effect. As Ruysen et al. (2014) pointed out, the immigration situation in OECD countries shows that immigrants from developing countries use public expenditure as a safety net, prefer countries with rising social expenditure, and show the magnet effect of public welfare. Third, the resident’s intention to live for more than five years does not mean permanent residence in the locality. Once they decide to leave the city after more than five years, the excessively high housing price has not become the main consideration in his migration decision. This characteristic of the city is shown as the migration neutral effect. In addition, the house price of individual small cities is not high. At this time, the house price will not become the key factor that affects the settlement or migration decision of the population mobility, which also belongs to the migration neutral effect. The above-mentioned three effects jointly determine the impact of housing prices on the China’s internal migration. For other city-level variables, the shares of books and doctors in total population (10,000 persons) in the city reflect that the higher the quality of public services to some extent, the more likely it is that the mobility population will live the focal place for a long time. While the population density in the city have no significant influence on the internal migration. Finally, higher removal rates of solid waste ($rsw$) and wastewater ($rww$) means the improvement of ambient air quality, which contributes to improve $wtl$. 

Table 2. Estimation results of the baseline models with dependent variable: wtl

| Variables | (1) Probit | (2) Probit | (3) Probit | (4) Logit | (5) CLog-Log | (6) OLS |
|-----------|------------|------------|------------|----------|--------------|--------|
| maqi      | 0.0003*    | 0.0010***  | 0.0008***  | 0.0015***| 0.0012***    | 0.0002***|
|           | (1.6931)   | (4.7403)   | (3.7157)   | (3.8311) | (3.7519)     | (3.5834) |
| fincome   | -0.1137*** | -0.0358*** | -0.0407*** | -0.0761***| -0.0683***   | -0.0082***|
|           | (-13.7536) | (-3.6230)  | (-4.0044)  | (-4.3727) | (-4.6765)    | (-3.6067) |
| hnature   | 0.1060***  | 0.0906***  | 0.1407***  | 0.1084*** | 0.0202***    |        |
|           | (5.9192)   | (4.9186)   | (4.3914)   | (4.0001) | (4.9542)     |        |
| age       | 0.0054***  | 0.0051***  | 0.0087***  | 0.0075*** | 0.0013***    |        |
|           | (7.2678)   | (6.6162)   | (6.5188)   | (6.5117) | (6.4821)     |        |
| gender    | 0.0681***  | 0.0644***  | 0.1075***  | 0.0850*** | 0.0151***    |        |
|           | (6.2358)   | (5.7524)   | (5.5695)   | (5.1915) | (5.4259)     |        |
| nation    | 0.0595***  | 0.0752***  | 0.1241***  | 0.1028*** | 0.0191***    |        |
|           | (2.9180)   | (3.4279)   | (3.2615)   | (3.1829) | (3.6189)     |        |
| spoint    | 0.2670***  | 0.2638***  | 0.4584***  | 0.3876*** | 0.0714***    |        |
|           | (24.4331)  | (23.7043)  | (23.9806)  | (24.2356) | (23.0255)    |        |
| leave     | -0.0318*** | -0.0321*** | -0.0599*** | -0.0544***| -0.0078***   |        |
|           | (-32.0934) | (-31.4978) | (-31.5332) | (-32.0774) | (-35.2972)   |        |
| house     | -1.3145*** | -1.3057*** | -2.7120*** | -2.5781***| -0.1758***   |        |
|           | (-50.1900) | (-48.9594) | (-42.6367) | (-41.2496) | (-81.5683)   |        |
| soul      | -0.2045*** | -0.1858*** | -0.3256*** | -0.2754***| -0.0402***   |        |
|           | (-13.1647) | (-11.6180) | (-11.6273) | (-11.3623) | (-11.3141)   |        |
| haf       | 0.0670***  | 0.0628**   | 0.1043**   | 0.0855**  | 0.0145***    |        |
|           | (2.8009)   | (2.5632)   | (2.4223)   | (2.3141)  | (2.9044)     |        |
| nai       | 0.0391***  | 0.0382***  | 0.0677***  | 0.0569*** | 0.0105***    |        |
|           | (3.4168)   | (3.2664)   | (3.3561)   | (3.3411)  | (3.5134)     |        |
| price     | -0.0169*   | -0.0351**  | -0.0330**  | -0.0030   |              |        |
|           | (-1.8566)  | (-2.2255)  | (-2.4860)  | (-1.4007) |              |        |
| popd      | 0.1989     | 0.3160     | 0.2914     | 0.0903*   |              |        |
|           | (0.9943)   | (0.9193)   | (1.0162)   | (1.8037)  |              |        |
| books     | -0.0003*** | -0.0006*** | -0.0005*** | -0.0001***|              |        |
|           | (-6.2420)  | (-6.1484)  | (-6.1612)  | (-6.1471) |              |        |
| doctors   | -0.0054*** | -0.0093*** | -0.0078*** | -0.0014***|              |        |
|           | (-11.2454) | (-11.4573) | (-11.8897) | (-10.7490) |              |        |
| rsw       | 0.0020***  | 0.0037**   | 0.0032***  | 0.0005*** |              |        |
|           | (6.2391)   | (6.5702)   | (6.6893)   | (6.3144)  |              |        |
| rswv      | 0.0006     | 0.0013     | 0.0011     | 0.0001    |              |        |
|           | (1.1877)   | (1.4846)   | (1.5608)   | (1.0053)  |              |        |
| elementary | -0.0542    | -0.0426    | -0.0783    | -0.0714   | -0.0085     |        |
|           | (-1.1981)  | (-0.8851)  | (-0.9413)  | (-1.0065) | (-0.6709)    |        |
| junior    | -0.0643    | -0.0457    | -0.0872    | -0.0790   | -0.0063     |        |

continued on next page
When analyzing the migration decision-making problem of the mobility caused by environmental pollution, tough the settlement decision of a single individual cannot affect the level of environmental pollution, the total effect of individual decision preference may affect the environmental pollution through various channels. For instance, when large-scale internal migration occurs, it may impose an indirect impact on environmental pollution by affecting employment and industrial structure. Qin and Liao (2016) examined whether rural-urban migration flows were associated with air pollution in China’s cities. They find that a strong negative association of in-migration with urban air quality. Furthermore, Rafiq et al. (2017) also found that inter-provincial migration has contributed to pollution. Given that, the simultaneous endogenous problem between the air pollution and willingness to leave may exist. To address the endogeneity issues, two-stage least squares (2SLS) regression analysis is adopted. Previous studies instrument air pollution with the strength of thermal inversions (Hicks et al., 2015; Arceo et al., 2016; Fu et al., 2017; Chen et al., 2017; Jans et al., 2018). However, due to the data limitations, historical air pollution index (API), spatial weighted value of PM2.5 in neighboring cities (Zheng et al., 2019) and topographic fluctuation of cities are treated as instrument variables (IVs). Given the existence of the IVs, the following two methods are used. In the first stage, a new variable is created using the IVs. In the second stage, the model-estimated values from stage one is then used in place of the actual values of the problematic

Table 2. Continued

| Variables    | (1)  | (2)  | (3)  | (4)  | (5)  | (6)  |
|--------------|------|------|------|------|------|------|
|              | Probit | Probit | Probit | Logit | CLog-Log | OLS |
| high         | (-1.4676) | (-0.9792) | (-1.0816) | (-1.1497) | (-0.5141) |
| associate    | (-0.1088**) | (-0.0937*) | (-0.1702**) | (-0.1528**) | (-0.0189) |
| bachelor     | (-2.4119) | (-1.9522) | (-2.0549) | (-2.1676) | (-1.5017) |
| master       | (-0.2320***) | (-0.2154***) | (-0.3973***) | (-0.3484***) | (-0.0489***) |
| first-marriage | (-4.8080) | (-4.2173) | (-4.5021) | (-4.6433) | (-3.7392) |
| remarriage   | (-0.2430***) | (-0.2411***) | (-0.4511***) | (-0.4020***) | (-0.0572***) |
| divorce      | (-1.7891) | (-1.6137) | (-1.9004) | (-2.0328) | (-1.9687) |
| widowed      | (-0.2647*) | (-0.2425) | (-0.5130*) | (-0.4738**) | (-0.0507**) |
| first-marriage | (-37.2717) | (-36.2608) | (-35.5211) | (-35.3612) | (-37.9294) |
| remarriage   | (-0.6710***) | (-0.6617***) | (-1.1460***) | (-0.9646***) | (-0.1954***) |
| divorce      | (-12.7655) | (-12.3361) | (-11.9621) | (-11.3606) | (-17.5742) |
| widowed      | (-0.1826**) | (-0.2265***) | (-0.3531***) | (-0.2784***) | (-0.0789***) |
| remarriage   | (-4.6152) | (-5.3398) | (-5.0579) | (-4.8859) | (-6.2760) |
| widowed      | (-0.2547***) | (-0.2528***) | (-0.4085***) | (-0.3262***) | (-0.0920***) |
| Constant     | -0.0523 | -0.0664 | 0.3850*** | 0.7855*** | 0.3920** | 0.5311*** |
| Observations | 89271  | 89271  | 86243  | 86243  | 86243  | 86243  |
| Log Likelihood | -45101.5800 | -39458.6900 | -37832.7000 | -37761.8700 | -37717.4100 | -38760.0600 |
| AIC          | 90209.1600 | 78963.7000 | 75723.4000 | 75581.7300 | 75492.8300 | 77578.1300 |
| BIC          | 90237.3600 | 79179.5600 | 75994.9900 | 75853.3100 | 75764.4100 | 77849.7100 |

Notes: 1) Average marginal effects are reported in the table (hereinafter); 2) robust z-statistics in parentheses; 3) *** p<0.01, ** p<0.05, * p<0.1; 4) the R-squared of OLS model is 0.1105.
predictors to compute Probit/OLS model for the response of interest. Due to space limitations, we report the second stage estimation results of 2SLS which show in Table 3. The results show the air pollution exerts significant and positive impacts on \( w_{t}\) and Kleibergen-Paap Wald rk F statistics are more than 10, thus the IVs selected in this study are relative valid.

5.2 Heterogeneity Analysis

On the one hand, we consider the migration-hierarchy-based heterogeneity analysis, i.e., the differential internal migration ranges, inter-province, intra-city within the same province and intra-county within the same city. Estimation results are shown in Table 4. The findings indicate that compared to inter-province and intra-city within the same province, residents are more willingness to leave in intra-county with the same city. One of the potential explanations is that with the migration scope expands, more extra implicit and explicit costs are required.

On the other hand, following Yu et al. (2019), we divide our sample into four groups, i.e., eastern cities vs. central/western cities; cities listed as the key environmental protection (KEP) vs. Non-KEP cities; cities listed as the resource-based (RB) and Non-RB; cities listed as two control zones (TCZ) and Non-TCZ. Estimation results of city-based heterogeneity analysis are summarized in Table 5. The results show that individuals from the cities in central and western regions, KEP cities and Non-RB have stronger willingness to leave than those from the cities in eastern region, Non-KEP cities and RB cities. Take Probit estimation results for example, regarding to magnitude, a 100 point increase in \( maqi \) would lead to a significantly increase in \( w_{t}\) by 13%, 16% and 15% for central/western, Non-KEP and RB cities, respectively. With the advanced development of economic and location advantage in eastern regions, it seems that residents from this area are less willingness to leave.

5.3 Robustness Checks

In this subsection, estimation results of different measures of air pollution are discussed. As shown in Table 6, we adopt the total days with AQI>200 (\( aqi200 \)), PM2.5 (\( pm25 \)) and SO\(_2\) (\( so2 \)) as three alternative measures and reexamine the impacts of air pollution on internal migration. The concentration of SO\(_2\) is reported in the data, while the concentration of PM2.5 is calculated following
Table 4. Estimation results of migration-hierarchy-based heterogeneity analysis

| Variables         | Probit       | Logit        | CLog-Log     | OLS          |
|-------------------|--------------|--------------|--------------|--------------|
|                   | Panel A: Inter-province |             |              |              |
| maqi              | 0.0007*      | 0.0012*      | 0.0009*      | 0.0002*      |
|                   | (1.7922)     | (1.8410)     | (1.8330)     | (1.9022)     |
| Individual FE     | Yes          | Yes          | Yes          | Yes          |
| City FE           | Yes          | Yes          | Yes          | Yes          |
| Log Likelihood    | -18694.2500  | -18670.0500  | -18651.6100  | -19563.6300  |
| AIC               | 37446.5000   | 37398.0900   | 37361.2200   | 39185.2700   |
| BIC               | 37694.2500   | 37645.8400   | 37608.9700   | 39433.0200   |
| Observations      | 37915        |              |              |              |
|                   | Panel B: Intra-city within the same province |             |              |              |
| maqi              | 0.0010**     | 0.0019***    | 0.0016***    | 0.0002**     |
|                   | (2.4944)     | (2.6291)     | (2.5896)     | (2.2244)     |
| Individual FE     | Yes          | Yes          | Yes          | Yes          |
| City FE           | Yes          | Yes          | Yes          | Yes          |
| Log Likelihood    | -12111.8500  | -12086.5300  | -12074.0800  | -11822.6600  |
| AIC               | 24281.7000   | 24231.0600   | 24206.1700   | 23703.3200   |
| BIC               | 24523.7700   | 24473.1300   | 24448.2400   | 23945.3900   |
| Observations      | 31173        |              |              |              |
|                   | Panel C: Intra-county within the same city |             |              |              |
| maqi              | 0.0013***    | 0.0022***    | 0.0019***    | 0.0003***    |
|                   | (3.0580)     | (3.0362)     | (2.9585)     | (2.9614)     |
| Individual FE     | Yes          | Yes          | Yes          | Yes          |
| City FE           | Yes          | Yes          | Yes          | Yes          |
| Log Likelihood    | -6808.4060   | -6789.8280   | -6778.4470   | -6714.9580   |
| AIC               | 13674.8100   | 13637.6600   | 13614.8900   | 13487.9200   |
| BIC               | 13899.5600   | 13862.4100   | 13839.6500   | 13712.6700   |
| Observations      | 17155        |              |              |              |

Notes: 1) Robust z-statistics in parentheses; 2) *** p<0.01, ** p<0.05, * p<0.1.
Table 5. Estimation results of city-based heterogeneity analysis.

| Variables | Probit | Logit | CLog-Log | OLS | Probit | Logit | CLog-Log | OLS |
|-----------|--------|-------|----------|-----|--------|-------|----------|-----|
|           | Eastern cities | Central/Western cities |           |     |        |       |          |     |
|          | maqi | -0.0001 | -0.0001 | <0.0001 | 0.0013*** | 0.0023*** | 0.0019*** | 0.0002*** |
|          |     | (-0.2226) | (-0.1796) | (-0.1390) | (0.4003) | (4.3164) | (4.4543) | (4.2983) |
|          | Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | City FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | Log Likelihood | -17142.5000 | -17113.1300 | -17092.8200 | -17648.4700 | -20569.5200 | -20533.0800 | -20513.1600 | -20921.5900 |
|          | AIC | 34343.0000 | 34284.2700 | 34243.6400 | 35354.9300 | 41197.0300 | 41124.1600 | 41084.3200 | 41901.1800 |
|          | BIC | 34590.5300 | 34531.8000 | 34491.1700 | 35602.4600 | 41451.9900 | 41379.1200 | 41339.2800 | 42156.1300 |
| Observations | 37627 | 48616 |           |     |        |       |          |     |

Panel B: KEP vs. Non-KEP

| Variables | Probit | Logit | CLog-Log | OLS | Probit | Logit | CLog-Log | OLS |
|-----------|--------|-------|----------|-----|--------|-------|----------|-----|
|           | KEP cities | Non-KEP cities |           |     |        |       |          |     |
|          | maqi | 0.0002 | 0.0003 | 0.0003 | 0.0001 | 0.0016*** | 0.0030*** | 0.0026*** | 0.0004*** |
|          |     | (0.4183) | (0.4698) | (0.4873) | (1.0871) | (5.0173) | (5.4031) | (5.7429) | (4.5890) |
|          | Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | City FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | Log Likelihood | -26141.6500 | -26097.3800 | -26071.2600 | -26599.7000 | -11564.6900 | -11534.3100 | -11515.1200 | -11992.3600 |
|          | AIC | 52341.3100 | 52252.7600 | 52200.5100 | 52357.3900 | 23187.3700 | 23126.6200 | 23088.2400 | 24042.7200 |
|          | BIC | 52602.4800 | 52531.9300 | 52461.6900 | 53518.5700 | 23424.1900 | 23363.4400 | 23325.0500 | 24279.5400 |
| Observations | 60236 | 26007 |           |     |        |       |          |     |

Panel C: RB vs. Non-RB

| Variables | Probit | Logit | CLog-Log | OLS | Probit | Logit | CLog-Log | OLS |
|-----------|--------|-------|----------|-----|--------|-------|----------|-----|
|           | RB cities | Non-RB cities |           |     |        |       |          |     |
|          | maqi | 0.0015*** | 0.0027*** | 0.0022*** | 0.0004*** | 0.0005* | 0.0009* | 0.0008* | 0.0001 |
|          |     | (3.9093) | (4.0331) | (3.9217) | (3.5155) | (1.6714) | (1.7981) | (1.7928) | (1.5959) |
|          | Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | City FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | Log Likelihood | -7404.9720 | -7390.5820 | -7378.5550 | -7705.7370 | -30434.3400 | -30284.6400 | -30248.7900 | -10971.8200 |
|          | AIC | 14867.9400 | 14839.1600 | 14815.1100 | 15469.4700 | 60744.6800 | 60627.2800 | 60555.5700 | 62001.6300 |
|          | BIC | 15091.9900 | 15063.2100 | 15039.1500 | 15693.5200 | 61010.0000 | 60892.6000 | 60820.8900 | 62266.9600 |
| Observations | 16741 | 69502 |           |     |        |       |          |     |

Panel D: TCZ vs. Non-TCZ

| Variables | Probit | Logit | CLog-Log | OLS | Probit | Logit | CLog-Log | OLS |
|-----------|--------|-------|----------|-----|--------|-------|----------|-----|
|           | TCZ cities | Non-TCZ cities |           |     |        |       |          |     |
|          | maqi | 0.0005 | 0.0008 | 0.0006 | 0.0001 | 0.0003 | 0.0005 | 0.0003 | 0.0001 |
|          |     | (1.5273) | (1.4639) | (1.3635) | (1.4664) | (0.7373) | (0.7467) | (0.5845) | (1.0893) |
|          | Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | City FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|          | Log Likelihood | -28093.7000 | -28037.7200 | -28001.5800 | -28714.3800 | -9568.0000 | -9552.0480 | -9542.9600 | -9835.2310 |
|          | AIC | 56245.4000 | 56133.4300 | 56061.1600 | 57486.7600 | 19194.0000 | 19162.1000 | 19143.9200 | 19728.4600 |
|          | BIC | 56508.0200 | 56396.0600 | 56323.7900 | 57749.3900 | 19427.1500 | 19395.2400 | 19377.0700 | 19961.6100 |
| Observations | 63327 | 22916 |           |     |        |       |          |     |

Notes: 1) Robust z-statistics in parentheses; 2) *** p<0.01, ** p<0.05, * p<0.1.
Buchard et al. (2016). Empirical results of these measures for various models indicate that air pollution exert significantly and positively impacts on willingness to leave. Thus, our empirical findings are robust to alternative measures of air pollution and models specification.

### 6. Conclusions and Policy Implications

#### 6.1 Conclusions

Air pollution has become a huge threat during sustainable development in the world. In the presence of air pollution, actual decision-making of the residents still needs to take into account the family

---

**Table 6. Alternative measures of air pollution.**

| Variables | Probit | Logit | CLog-Log | OLS |
|-----------|--------|-------|----------|-----|
| **Panel A: Air pollution measured by total days with AQI>200** |
| aqi200 | 0.0006** | 0.0010** | 0.0009** | 0.0002*** |
| (2.1695) | (2.1523) | (2.2555) | (2.9335) |
| Individual FE | Yes | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes | Yes |
| Log Likelihood | -39467.7200 | -39400.4100 | -39360.5300 | -40498.5900 |
| AIC | 78981.4400 | 78846.8300 | 78767.0700 | 81043.1700 |
| BIC | 79197.6300 | 79063.0100 | 78983.2600 | 81259.3600 |
| Observations | 89271 |
| **Panel B: Air pollution measured by PM2.5** |
| pm25 | 0.0009* | 0.0017** | 0.0017** | 0.0003** |
| (1.8652) | (2.1016) | (2.4266) | (2.3283) |
| Individual FE | Yes | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes | Yes |
| Log Likelihood | -37790.2500 | -37719.5300 | -37674.1100 | -38716.2400 |
| AIC | 75638.5100 | 75497.0600 | 75406.2100 | 77490.4700 |
| BIC | 75910.0500 | 75768.6000 | 75677.7500 | 77762.0100 |
| Observations | 86120 |
| **Panel C: Air pollution measured by SO2** |
| so2 | 0.0013*** | 0.0022*** | 0.0019*** | 0.0003*** |
| (3.0580) | (3.0362) | (2.9585) | (2.9614) |
| Individual FE | Yes | Yes | Yes | Yes |
| City FE | Yes | Yes | Yes | Yes |
| Log Likelihood | -37832.4800 | -37760.5200 | -37714.1000 | -38759.2100 |
| AIC | 75722.9700 | 75579.0500 | 75486.2100 | 77576.4300 |
| BIC | 75994.5500 | 75850.6300 | 75757.7900 | 77848.0100 |
| Observations | 86243 |

Notes: 1) Robust z-statistics in parentheses; 2) *** p<0.01, ** p<0.05, * p<0.1.
income, housing attributes and prices, and other potential factors. In this study, we first match the data from Migrant Population Dynamic Monitoring Survey in China and average daily Air Quality Index (AQI) in 2014 at the city level and empirically examine the impact of air pollution on residents’ willingness to leave. Our findings show that, on average, the probability of the willingness to leave will grow by approximately 10% if the average value of AQI is increased by 100 points, *ceteris paribus*. Having addressed the heterogeneity issue, the effects are still holding. In particular, compared to inter-province and intra-city within the same province, residents are more willing to leave in intra-county with the same city. Besides, individuals from the cities in central and western regions, the listed key environmental protection (KEP) cities and cities not listed as resource-based (RB) cities have stronger willingness to leave than those from the cities in eastern region, Non-KEP cities and RB cities. The empirical results are robust to alternative measures of air pollution and models specification.

### 6.2 Policy Implications

To create a livable living environment for residents requires the government to strengthen environmental control, curb the deterioration of ecological environment and gradually improve environmental quality. The most direct and urgent way to control environmental pollution is to control the source of pollution and reduce the generation of pollution. There are many sources of environmental pollution, including industrial pollution emissions, automobile exhaust, domestic waste, etc. According to different sources of pollution, appropriate control measures should be taken to comprehensively reduce the generation of pollution and improve the quality of ecological environment. Moreover, heterogeneity analysis suggests that environmental protection policies should be formulated according to different city locations and city sizes. For example, in the central and western regions where environmental quality is more serious, air pollution has a greater impact on population mobility.

Industrial pollution emission is one of the main factors that cause environmental pollution in various regions of our country, and it is also an important link in controlling environmental pollution in various regions. At the beginning of the reform and opening up, due to the insufficient attention to environmental issues, one-sided pursuit of economic growth, coupled with fierce competition among local governments in attracting investment, the situation of sacrificing the ecological environment for economic development was formed, and the ecological environment gradually deteriorated. However, for the manufacturers, the pollution discharge behavior of the manufacturers results in obvious negative externalities. From the point of view of profit maximization, the manufacturers also have the motivation to increase the pollution discharge in order to reduce the production cost. Although these polluting enterprises have increased Gross Domestic Product in the short term and provided some jobs for the local people, in the long run they have done nothing but harm to a region’s sustainable economic and social development. All regions need to show courage in breaking their wrists, actively and steadily eliminate backward production capacity and dissolve excess production capacity, especially for industries prone to environmental pollution, such as steel industry, extractive industry, paper industry and chemical industry, and comprehensively use various administrative, judicial, economic and other means. A linkage mechanism for supervision and law enforcement should be established to promote the transformation and upgrading of polluting enterprises in the region, save energy and reduce consumption. Polluting enterprises that cannot be transformed can be withdrawn. Only in this way can they jointly promote green development and return the common people to a green landscape.

Automobile exhaust is also an important cause of environmental pollution, especially air pollution, which can be reduced from two aspects: automobile fuel and automobile maintenance. In terms of automobile fuel, on the one hand, it is necessary to actively promote the improvement of oil quality. According to the requirements of the national air pollution control, fuel quality standards should be defined and formulated, and such standards should be dynamic. With the passage of time and the improvement of technical level, the standards will also be improved. In the nationwide comprehensive
supply of four national standards of automobile gasoline and diesel, conditional areas can take the lead in the supply of five national standards of automobile gasoline and diesel, and gradually to the national promotion. On the other hand, it is necessary to further promote the use of electric vehicles, Hydrogen vehicle and other zero emission or ultra-low emission cars, and accelerate the research and development and commercial application of new automobile fuels.

In addition to agricultural pollution, industrial pollution and automobile exhaust, the increasing total energy consumption and the energy consumption structure dominated by media are also important causes of environmental pollution. In 2015, China’s total energy consumption has reached 4.3 billion tons of standard coal, of which coal consumption accounts for 64% of the total energy consumption. This energy utilization has brought great pressure to China’s environmental pollution. Therefore, energy utilization efficiency must be improved. To reduce waste in the process of energy utilization, improve the structure of energy consumption, actively promote clean and efficient use of coal resources, and ensure the continuous negative growth of coal consumption in Beijing, Tianjin, Hebei and other regions. Promote the safe use of renewable energy sources such as solar energy, wind energy and tidal energy, and improve the recovery and utilization rate of secondary energy. Popularize and promote energy-saving and emission-reduction technologies, expand the utilization of desulfurization equipment in coal motor units, steel, cement and other production lines, strengthen the combination of production, study and research, and promote the research and development and upgrading of key energy-saving technologies.

Furthermore, governors should formulate corresponding environmental protection and regulation policies according to different stages of urban development in the presence of economic and social development. For example, the national carbon emissions trading market launched online trading on July 16, 2021 and the power generation industry was the first industry to be included, with 2,225 power companies taking the lead. Thus, policy makers should seize the opportunity of carbon emission trading pilot market and finally achieve greener development. Meanwhile, to accelerate the establishment of a “dual circulation” development pattern in which domestic economic cycle plays a leading role while international economic cycle remains its extension and supplement, and in the process of urban development, restrictions on population mobility should be relaxed, especially household registration system (Hukou) in China’s actual situation.

Conflicts of Interest
The authors declare no conflict of interest.

FUNDING AGENCY
Open Access Funding for this article has been covered by the authors of this manuscript.

ACKNOWLEDGEMENTS
The authors would like to thank the financial support provided by the National Natural Science Foundation of China (71903068) and the Shanghai High-level University Development Project.
REFERENCES

Alexeev, A., Good, D. H., & Reuveny, R. (Eds.). (2011). *Climate change, weather disasters and international migration*. EAERE 18th Annual Conference, Special session on Environmentally Induced Migration, Rome, Italy.

Arceo, E., Hanna, R., & Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. *Economic Journal (London)*, 126(591), 257–280. doi:10.1111/ecoj.12273

Bates, D. (2002). Environmental refugees? Classifying human migrations caused by environmental change. *Population and Environment*, 23(5), 465–477. doi:10.1023/A:1015186001919

Beegle, K., De Weerdt, J., & Dercon, S. (2011). Migration and economic mobility in Tanzania: Evidence from a tracking survey. *The Review of Economics and Statistics*, 93(3), 1010–1033. doi:10.1162/REST_a_00105

Beine, M., & Parsons, C. (2015). Climatic Factors as Determinants of International Migration. *The Scandinavian Journal of Economics*, 117(2), 723–767. doi:10.1111/sjoe.12098

Black, R., Adger, W. N., Arnell, N. W., Dercon, S., Geddes, A., & Thomas, D. (2011). The effect of environmental change on human migration. *Global Environmental Change, 21*, S3–S11. doi:10.1016/j.gloenvcha.2011.10.001

Bohra-Mishra, P., Oppenheimer, M., & Hsiang, S. M. (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences of the United States of America*, 111(27), 9780–9785. doi:10.1073/pnas.1317166111 PMID:24958887

Borjas, G. J. (1989). Economic theory and international migration. *The International Migration Review*, 23(3), 457–485. doi:10.1177/019791838902300304 PMID:12282789

Boustan, L., Kahn, M., & Rhode, P. (2012). Moving to higher ground: Migration response to natural disasters in the early twentieth century. *The American Economic Review*, 102(3), 238–244. doi:10.1257/aer.102.3.238 PMID:25067842

Buchard, V., da Silva, A. M., Randles, C. A., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., & Winker, D. (2016). Evaluation of the surface PM2.5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States. *Atmospheric Environment*, 125, 100–111. doi:10.1016/j.atmosenv.2015.11.004

Carvajal, L., & Medalho Pereira, I. (2009). Climate Shocks and Human Mobility: Evidence from Nicaragua. SSRN Electronic Journal. 10.2139/ssrn.1599667

Chang, T. Y., Huang, W., & Wang, Y. (2018). Something in the air: Pollution and the demand for health insurance. *The Review of Economic Studies*, 85(3), 1609–1634. doi:10.1093/restud/rdy016

Chen, J., Gao, M., Mangla, S. K., Song, M., & Wen, J. (2020c). Effects of technological changes on China’s carbon emissions. *Technological Forecasting and Social Change*, 153, 119938. doi:10.1016/j.techfore.2020.119938

Chen, J., Wang, B., Huang, S., & Song, M. (2020a). The influence of increased population density in China on air pollution. *The Science of the Total Environment*, 735, 139456. doi:10.1016/j.scitotenv.2020.139456 PMID:32470669

Chen, J., Xu, C., Xie, Q., & Song, M. (2020b). Net primary productivity-based factors of China’s carbon intensity: A regional perspective. *Growth and Change*, 51(4), 1727–1748. doi:10.1111/grow.12423

Chen, S., Oliva, P., & Zhang, P. (2017). *The effect of air pollution on migration: evidence from China*. National Bureau of Economic Research. doi:10.3386/w24036

Chen, Z., Wang, J. N., Ma, G. X., & Zhang, Y.-S. (2013). China tackles the health effects of air pollution. *Lancet*, 382(9909), 1959–1960. doi:10.1016/S0140-6736(13)62064-4 PMID:24332069

Di Falco, S., Veronesi, M., & Yesuf, M. (2011). Does Adaptation to Climate Change Provide Food Security? A Micro-Perspective from Ethiopia. *American Journal of Agricultural Economics*, 93(3), 825–842. doi:10.1093/ajae/aar006

El-Hinnawi, E. (1985). *Environmental refugees*. UNEP.
Fisher, B., Turner, R. K., & Morling, P. (2009). Defining and classifying ecosystem services for decision making. *Ecological Economics, 68*(3), 643–653. doi:10.1016/j.ecolecon.2008.09.014

Fu, S., Viard, V. B., & Zhang, P. (2017). *Air Pollution and Manufacturing Firm Productivity: Nationwide Estimates for China*. Working Paper.

Gawande, K., Bohara, A. K., Berrens, R. P., & Wang, P. (2000). Internal migration and the environmental Kuznets curve for US hazardous waste sites. *Ecological Economics, 33*(1), 151–166. doi:10.1016/S0921-8009(99)00132-9

Gray, C. L. (2009). Environment, Land, and Rural Out-migration in the Southern Ecuadorian Andes. *World Development, 37*(2), 457–468. doi:10.1016/j.worlddev.2008.05.004

Gray, C. L. (2011). Soil quality and human migration in Kenya and Uganda. *Global Environmental Change, 21*(2), 421–430. doi:10.1016/j.gloenvcha.2011.02.004 PMID:22016577

Gray, C. L., & Mueller, V. (2012). Natural disasters and population mobility in Bangladesh. *Proceedings of the National Academy of Sciences of the United States of America, 109*(16), 6000–6005. doi:10.1073/pnas.1115944109 PMID:22474361

Gutmann, M. P., Deane, G. D., Lauster, N., & Peri, A. (2005). Two population-environment regimes in the Great Plains of the United States, 1930–1990. *Population and Environment, 27*(2), 191–225. doi:10.1007/s11111-006-0016-3

Halliday, T. (2006). Migration, Risk, and Liquidity Constraints in El Salvador. *Economic Development and Cultural Change, 54*(4), 893–925. doi:10.1086/503584

Hao, Y., Hao, X. Y., Li, Y. W., Zhang, Y. R., & Wu, H. T. (2020, June). How does air quality affect the willingness of graduate students to stay? Evidence from Beijing city, China. *Journal of Cleaner Production, 259*, 120759. doi:10.1016/j.jclepro.2020.120759

Hassani-Mahmooei, B., & Parris, B. W. (2012). Climate change and internal migration patterns in Bangladesh: An agent-based model. *Environment and Development Economics, 17*(6), 763–780. doi:10.1017/S1355770X12000290

Hicks, D., Marsh, P., & Oliva, P. (2015). *Air pollution and procyclical mortality: Causal evidence from thermal inversions*. Working Paper.

Jans, J., Johansson, P., & Nilsson, J. P. (2018). Economic status, air quality, and child health: Evidence from inversion episodes. *Journal of Health Economics, 61*, 220–232. doi:10.1016/j.jhealeco.2018.08.002 PMID:30193188

Kahn, M. E. (2000). Smog reduction’s impact on California county growth. *Journal of Regional Science, 40*(3), 565–582. doi:10.1111/0022-4146.00188

Koubi, V., Spilker, G., Schaffer, L., & Böhmelt, T. (2016). The role of environmental perceptions in migration decision-making: Evidence from both migrants and non-migrants in five developing countries. *Population and Environment, 38*(2), 134–163. doi:10.1007/s11111-016-0258-7

Lai, W. Y., Song, H., Wang, C., & Wang, H. H. (2021). Air pollution and brain drain: Evidence from college graduates in China. *China Economic Review, 68*, 101624. doi:10.1016/j.cieco.2021.101624

Lee, E. (1966). A Theory of Migration. *Demography, 3*(1), 47–57. doi:10.2307/2060063

Lim, Y. H., Kim, H., Kim, J. H., Bae, S., Park, H. Y., & Hong, Y.-C. (2012). Air pollution and symptoms of depression in elderly adults. *Environmental Health Perspectives, 120*(7), 1023–1028. doi:10.1289/ehp.1104100 PMID:22514209

Liu, Z. M., & Yu, L. (2021). Stay or Leave? The Role of Air Pollution in Urban Migration Choices. *Ecological Economics, 177*, 106780. doi:10.1016/j.ecolecon.2020.106780

Mbaye, L., & Zimmermann, K. F. (2016). Natural Disasters and Human Mobility. *Social Science Electronic Publishing, 10*(1), 37–56.

Millock, K. (2015). Migration and Environment. *Annual Review of Resource Economics, 7*(1), 35–60. doi:10.1146/annurev-resource-100814-125031
Myers, N. (1997). Environmental refugees. *Population and Environment, 19*(2), 167–182. doi:10.1023/A:1024623431924

Paul, B. K. (2005). Evidence against disaster-induced migration: The 2004 tornado in north-central Bangladesh. *Disasters, 29*(4), 370–385. doi:10.1111/j.0361-3666.2005.00298.x PMID:16277646

Perch-Nielsen, S. L., Bättig, M. B., & Imboden, D. (2008). Exploring the link between climate change and migration. *Climatic Change, 91*(3-4), 375. doi:10.1007/s10584-008-9416-y

Piguet, E. (2010). Linking climate change, environmental degradation, and migration: A methodological overview. *Wiley Interdisciplinary Reviews: Climate Change, 1*(4), 517–524. doi:10.1002/wcc.54

Qin, H., & Liao, T. F. (2016). The association between rural–urban migration flows and urban air quality in China. *Regional Environmental Change, 16*(5), 1375–1387. doi:10.1007/s10113-015-0865-3

Qin, Y., & Zhu, H. (2018). Run away? Air pollution and emigration interests in China. *Journal of Population Economics, 31*(1), 235–266. doi:10.1007/s00148-017-0653-0

Rafiq, S., Nielsen, I., & Smyth, R. (2017). Effect of internal migration on the environment in China. *Energy Economics, 64*, 31–44. doi:10.1016/j.eneco.2017.03.009

Rafiq, S., Salim, R., & Nielsen, I. (2016). Urbanization, openness, emissions, and energy intensity: A study of increasingly urbanized emerging economies. *Energy Economics, 56*, 20–28. doi:10.1016/j.eneco.2016.02.007

Reuveny, R. (2007). Climate change-induced migration and violent conflict. *Political Geography, 26*(6), 656–673. doi:10.1016/j.polgeo.2007.05.001

Reuveny, R., & Moore, W. H. (2009). Does Environmental Degradation Influence Migration? Emigration to Developed Countries in the Late 1980s and 1990s. *Social Science Quarterly, 90*(3), 461–479. doi:10.1111/j.1540-6237.2009.00569.x

Ruyssen, I., Everaert, G., & Rayp, G. (2014). Determinants and dynamics of migration to OECD countries in a three-dimensional panel framework. *Empirical Economics, 46*(1), 175–197. doi:10.1007/s00181-012-0674-1

Shen, H., Tao, S., Chen, Y., Ciais, P., Güneralp, B., Ru, M., Zhong, Q., Yun, X., Zhu, X., Huang, T., Tao, W., Chen, Y., Li, B., Wang, X., Liu, W., Liu, J., & Zhao, S. (2017). Urbanization-induced population migration has reduced ambient PM2.5 concentrations in China. *Science Advances, 3*(7), e1700300. doi:10.1126/sciadv.1700300 PMID:28776030

Todaro, M. P. (1969). A model of labor migration and urban unemployment in less developed countries. *The American Economic Review, 59*(1), 138–148.

Xu, P., Chen, Y., & Ye, X. (2013). Haze, air pollution, and health in China. *Lancet, 382*(9910), 2067–2067. doi:10.1016/S0140-6736(13)62693-8 PMID:24360386

Yang, G., Wang, Y., Zeng, Y., Gao, G. F., Liang, X., Zhou, M., Wan, X., Yu, S., Jiang, Y., Naghavi, M., Vos, T., Wang, H., Lopez, A. D., & Murray, C. J. L. (2013). Rapid health transition in China, 1990-2010: Findings from the Global Burden of Disease Study 2010. *Lancet, 381*(9882), 1987–2015. doi:10.1016/S0140-6736(13)61097-1 PMID:23746901

Yu, Y., Hu, H., Zhang, Y., & Yin, Z. (2019). Metafrontier Eco-Efficiency and Its Convergence Analysis for China: A Multidimensional Heterogeneity Perspective. *Emerging Markets Finance & Trade, 55*(7), 1531–1549. doi:10.1080/1540496X.2018.1559142

Zheng, S., Wang, J., Sun, C., Zhang, X., & Kahn, M. E. (2019). Air pollution lowers Chinese urbanites’ expressed happiness on social media. *Nature Human Behaviour, 3*(3), 237–243. doi:10.1038/s41562-018-0521-2 PMID:30953012
Yantuan Yu is an associate research fellow in Guangdong University of Foreign Studies, China. His research interests are Environmental Economics, Energy Economics, Applied Economics, Spatial Economics, Efficiency and Productivity Analysis. Dr. Yu has published papers in Energy Economics, Energy Policy, Journal of Environmental Management, Ecological Modelling, etc.

Zhao Weidong is a professor at School of Finance, Shanghai Lixin University of Accounting and Finance, China.

Xiuzhen Li is an associate Professor at School of International Economy and Trade, Shanghai Lixin University of Accounting and Finance, China. Her research interests are Environmental Economics, Applied Economics, International Trade, Data Finance. And she is the co-corresponding author of this paper. Dr. Li has published more than 30 academic papers in many academic journals, including Emerging Markets Finance and Trade, Economic Modelling, Energies, etc.

Yun Zhang is a professor at School of Finance, Shanghai Lixin University of Accounting and Finance, China. His research interests are Environmental Economics, Applied Economics, Green Finance, Data Finance. And he is the co-corresponding author of this paper. Dr. Zhang has published more than 70 academic papers in many academic journals, including Environmental Science and Pollution Research, Pacific-Basin Finance Journal, European Journal of Finance, World Economy, Economic Modelling, Finance Research Letters, etc.

Xiao Miao is a professor at Innovation Research Institute of T.C.M, Shanghai University of T.C.M, Shanghai, China. Dr. Miao is a board-certificated dermatologist and physician scientist. She is also interested in the study of big health data through bioinformatics and machine learning approaches. And she is the co-corresponding author of this paper. Dr. Miao has published 20 SCI papers in high-impact journals, including Nature Communication, International Journal of Dermatology, Cell Death Discovery, Emerging Markets Finance and Trade, Energies, etc.