The Productivity Consequences of Pollution-Induced Migration in China

Gaurav Khanna*
Wenquan Liang†
A. Mushfiq Mobarak‡
Ran Song §

October, 2019

Results are preliminary and subject to change

Abstract

Migration and pollution are two of the defining features of China's impressive growth performance over the last 30 years. In this paper we study the migration response to the dispersion of pollution across Chinese cities, and its consequences for productivity and welfare. We document a robust pattern in which college-educated workers emigrate more in response to pollution than the unskilled. Their greater sensitivity to air quality holds up in cross-sectional variation across cities, panel variation with city fixed-effects, in a regression discontinuity, and when instrumenting for pollution using power-plants upwind of cities, or with thermal inversions that trap pollution. Pollution therefore changes the spatial distribution of skilled and unskilled workers, which results in higher returns to skill in cities that the educated migrate away from. We quantify the loss in aggregate productivity due to this sorting by estimating a model of demand and supply of skilled and unskilled workers across Chinese cities. Reducing pollution increases GDP both by directly improving health and productivity, and indirectly by changing the spatial distribution of skilled and unskilled workers. Counterfactual simulations show that gains through the indirect spatial sorting channel are about as large as the direct health benefits of clean air. *Hukou* policy restrictions on mobility exacerbates the losses in productivity and welfare.

Keywords: Internal migration, air pollution, spatial productivity gap
JEL Codes: Q52, R12, J61

*University of California – San Diego, gakhanna@ucsd.edu
†Jinan University, liang.wenquan@jnu.edu.cn
‡Yale University, NBER, Deakin Univ, and CEPR, ahmed.mobarak@yale.edu
§Yale University, ran.song@yale.edu
We thank seminar participants at UCSD, UW (Seattle), NBER Environment and Energy Summer Institute, U Richmond, NBER Urban Economics Summer Institute, Stanford SITE Migration Conference, Cities and Development Conference (Johns Hopkins), Ortygia Business School, Harvard, CU Hong Kong, and Lingnan U for useful comments, and to Ed Glaeser, Matthew Kahn, Paulina Oliva and David Yang for insightful suggestions.

1 Introduction

The large productivity gaps across regions or sectors within developing countries (Restuccia and Rogerson, 2013, 2017) create an enduring development puzzle: Why do workers remain in low productivity areas when they could experience wage gains elsewhere (Gollin et al., 2014)? It is important to understand the drivers of worker location choices, because spatial reallocation has the potential to produce substantial productivity gains (Bryan et al., 2014; Clemens et al., 2019). The literature has proposed a few explanations for the low rates of within-country mobility observed in the world: Migration costs (Bazzi, 2017; Bryan and Morten, 2018), income risk at destinations (Bryan et al., 2014; Lewis, 1954), non-transferable location-specific amenities (Munshi and Rosenzweig, 2016), or other urban disamenities (Lagakos et al., 2019), like pollution (Heblich et al., 2019).

Pollution can have a large effect on where people choose to live. Indian human resource agencies report that skilled workers are now looking to relocate away from polluted Delhi to cleaner cities like Pune (Sharma and Chandna, 2019). Particulate matter pollution exceeds WHO air quality guidelines for 96% of Chinese cities in 2015, and on average four times higher than the level considered safe. Chinese air pollution reduces citizens' life expectancy and causes elevated rates of heart disease, stroke, and lung cancer (Ebenstein et al., 2015, 2017). We document that the sharp increases in pollution in China in recent years were concentrated in a few cities, which increased the cross-city dispersion in pollution, and with it, incentives to emigrate. Increased pollution and migration have thus been two of the defining features of the impressive Chinese growth experience over the last 30 years (Brandt et al., 2008; Tombe and Zhu, 2019; Zheng and Kahn, 2013).

This paper empirically analyzes whether workers relocate in response to variation in air quality across Chinese cities, and then quantifies the aggregate productivity consequences of this movement. While one branch of the literature argues that workers are efficiently sorted (Young, 2014), we show that asymmetric migration responses of skilled and unskilled workers to pollution can create losses in aggregate productivity. This is because skilled workers choose to leave polluted places where they would be more productive, and the production function complementarity between skilled and unskilled workers also makes the unskilled less productive when skilled people leave. Chinese *Hukou* policy exacerbates those losses and reduces welfare of the poor, because it differentially restricts mobility of unskilled workers.

Pollution and the skill-composition of the workforce are jointly determined, and both depend on other factors such as industrial sector growth. The first part of our paper is therefore careful to identify the migration response to *exogenous variation* in pollution. To build confidence that our estimates indeed represent the causal effect of air quality on mobility, we assemble several different datasets, and investigate this relationship under multiple independent sources of data variation. We isolate exogenous fluctuations in pollution leveraging variation in wind direction and the historical placement of thermal power plants (as in Freeman et al., 2019), a regression discontinuity around the Huai river (as in Chen et al., 2013), and a meteorological phenomenon called thermal inversions that traps pollution (as in Arceo et al., 2016; Chen et al., 2017; Hicks et al., 2015). We also model changes in migration as a function of changes in pollution over time, exploiting the panel dimension of our data. Across all these research designs, we find robust evidence that college-educated workers move to areas with lower levels of pollution, while the less educated are comparatively less responsive.

Quantifying the productivity implications of this mobility response requires a model, because the differential emigration of skilled workers changes the skilled wage premia across cities, which in turn also affects the location choices of the unskilled in general equilibrium. We empirically document that the relative scarcity of skilled workers in polluted cities raises the marginal product of skill in those locations. Cleaning up polluted cities therefore induces a relocation of skilled workers from low marginal product areas to high marginal product areas, which raises aggregate output, as in Hsieh and Klenow (2009) and Hsieh and Moretti (2018). We also document that skilled and unskilled workers are complements in production, which implies that the in-migration of skilled workers makes the unskilled they join more productive.

The estimated model allows us to quantify the magnitudes of these productivity shifts. The differential response to (exogenous variation in) pollution by skill group shifts the labor supply of workers, and produces a valid estimate of the compensating wage-differential that skilled workers have to be paid to reside in polluted cities. Next we use the permanent normalization of trading relations (PNTR) between the US and China, which generated variation in the demand for skilled and unskilled workers across Chinese cities, to trace out the relative labor supply curve, and produce a valid estimate of the labor supply response to changes in wages.[1]

Our quantification exercise is complicated by the facts that pollution directly affects health and productivity (Zivin and Neidell, 2012); production, in turn, affects air quality (Andreoni and Levinson, 2001); and worker location decisions may affect agglomeration (Au and Henderson, 2006), house prices (Bayer et al., 2009), or the pollution-intensity of production. We incorporate all these mechanisms in our model, introducing additional instruments to estimate the relevant elasticities.[2] In summary, we quantify the productivity effects of pollution via worker re-sorting using our model, *accounting for* other important mechanisms through which production, pollution, productivity, and migration are related.

Estimating this model allow us to quantify how much of the wage gap across Chinese cities is attributable to pollution differences. Our estimates imply that equalizing pollution between high-pollution Beijing and low-pollution Ziyun would bridge the between-city skilled wage gap by 5.5%.[3] The fact that pollution explains a meaningful portion of the productivity gaps across cities sheds some light on the behavioral puzzle we raised at the outset, as to why workers remain in low-wage areas within countries. Survey data from India supports this explanation: When 9,000 Delhi residents were asked about their plans to deal with pollution, their single-most common response was "relocate" (Kapur, 2019).

To quantify the productivity loss from pollution, we use the model to perform a counterfactual exercise where we move pollution away from cities with more skill-biased capital to cities with less skill-biased capital. This is precisely the type of pollution-control program that the Chinese government has recently introduced. The *12th Five-Year Plan on Air Pollution*

[1] Pierce and Schott (2016) uses PNTR as an "import" shock to document effects on the United States, and it is simultaneously an "export" shock that had differential effects on skilled and unskilled labor demand in Chinese cities that were more or less exposed to trade with the US. With unique city-level data on the production of each product for which we have tariff information, we are able to construct a shift-share instrument.

[2] For instance, to estimate skilled-worker agglomeration effects, we leverage a large-scale university expansion in China at the turn of the century that rapidly expanded college enrollment by 20% in certain cities. These elasticities we estimate to close the model are similar to credible estimates in the literature on the direct effect of pollution on productivity (Adhvaryu et al., 2016; Chang et al., 2019), or of worker location on agglomeration (Zhang and Yao, 2010). As such, disciplining our model by borrowing relevant elasticities from the literature (instead of estimating them ourselves) produces similar quantitative results.

[3] Companies in China reportedly offer up to 20% wage premiums to induce workers to relocate to polluted Beijing, so our estimates appear to be in line with the real-world behavior of firms and workers. See, "*Asia's pollution exodus: Firms struggle to woo top talent*" https://phys.org/news/2019-03-asia-pollution-exodus-firms-struggle.html

Prevention and Control in Key Regions sets targets for ambient concentrations of particulate matter, with more stringent targets for high-productivity, polluted regions like Beijing-Tianjin.[4] This exercise increases aggregate GDP in China by about 3.86%, simply through the labor reallocation channel. As a benchmark, this is about as large as the direct effect of air pollution on worker health and labor productivity, as estimated in our model. The relationship between pollution and health has been the subject of a much larger literature in economics and epidemiology, but we learn that ignoring labor mobility grossly underestimates the overall consequences of air pollution on an economy's prosperity.

Whether relocating pollution also affects aggregate welfare (beyond productivity effects) depends on the precise underlying reason as to why the high and low-skilled react differently to pollution. Survey data shows that this is partly due to different preferences of the rich and environmentally-aware. Administrative records indicate that the skilled and unskilled also face vastly different migration costs under China's *hukou* system. Several Chinese cities have adopted a point-based system that exempts workers with skills or higher education from their hukou restrictions (see Appendix Table B2). Without the exemption, the system imposes a burden on poor in-migrants to cities by limiting or prohibiting their access to many government-provided benefits (Combes et al., 2019). When mobility is restricted, unskilled workers may be trapped in polluted cities with low wages even as their skilled counterparts leave. Relocating pollution away from such cities may raise welfare.

Once we incorporate the *hukou* system into our analysis using a city-level index of mobility restrictions, our model shows that the productivity losses from pollution are magnified in cities whose *hukou* policies are more restrictive. When unskilled workers cannot easily leave with their skilled counterparts, *hukou* restrictions exacerbate the mismatch between where workers are situated. Relocating pollution away from cities with skill-biased capital *and* relaxing hukou policy simultaneously raises GDP by as much as 18.8%.[5]

[4] http://www.mep.gov.cn/gkml/hbb/bwj/201212/t20121205_243271.htm, accessed September 17, 2019

[5] While China's hukou policy is unique, institutional restrictions on migration are not without precedent. For example, state-level entitlement schemes in India discriminate against out-of-state migrants and inhibit inter-state mobility (Kone et al., 2018). Furthermore, migration costs are high for the poor in most developing countries (Bryan and Morten, 2018). Transportation infrastructure is often of poor quality, and this poses a disproportionate burden on the poor. As industrialization in many developing countries worsens environmental quality, our results suggest that high migration costs can exacerbate the welfare and productivity losses from pollution.

Other research documents Chinese households' willingness to pay to avoid pollution using variation in the prices of housing (Freeman et al., 2019) and air filters (Ito and Zhang, 2019). The rich and educated are willing to pay more, similar to the emigration patterns we see. Firms in China pay substantial 'pollution premiums' to attract workers.[6] Most closely related to our empirical finding, Chen et al. (2017) also report that workers migrate in response to air quality. They infer this from data on population changes and find large mobility responses to pollution even during a period when information on air quality was not readily available.[7] The first part of our paper uses restricted-access data on actual migration decisions from 2000-2015 (before and after information about pollution was widely disseminated) to estimate this mobility response, and finds magnitudes similar to the Chen et al. (2017) estimates, but using very different data sources.

We describe our data sources in Section 2, describe geographic and time-series patters on pollution and migration in section 3, identification strategies in Section 4 and close the empirical part of our paper with estimates of the causal effect of pollution on migration in section 5. The quantitative part of the paper consists of the theoretical framework in Section 6, estimating model parameters in Section 7, and conducting counterfactual exercises in Section 9. Section 10 concludes.

2 Data

2.1 Demographic and Migration Data

We measure internal migration using the 2005, 2010 and 2015 Population Census of China. The 2015 Census is the latest census with restricted public access. Importantly, it is the only population census after both the 2008 disclosure of PM2.5 data by the US Embassy in China and the publication of PM2.5 data for Chinese cities by Chinese Government in 2012. The censuses record demographic and economic characteristics of individuals, including age,

[6] See, for instance, "*Asia's pollution exodus: Firms struggle to woo top talent*" https://phys.org/news/2019-03-asia-pollution-exodus-firms-struggle.html, and "*Companies in South China See Opportunity in Beijing's Smog*" https://www.nytimes.com/2015/12/23/world/asia/beijing-air-pollution-china-smog.html

[7] The US embassy started disclosing PM 2.5 data in Beijing in 2008. Towards the end of 2012, the Chinese government started releasing data more widely, and by 2013 most cities had publicly available PM 2.5 data.

gender, education level, employment status, occupation, *hukou* location and type (rural or urban), and current residential city. We combine the 2015 One-Percent Census sample with the 2005 One-Percent Census and the 2010 National Population Census. We use ages between 25 and 54 for our analysis across all three census waves.

We define migration in a few different ways. First, in the Census data, migrants are defined as those who are away from their *hukou* city for more than six months.[8] *Hukou* status in China determines citizens' access to state-provided goods (such as schools for their children) and services (such as marriage registries or passport renewals). *Hukou* status is therefore a strong indicator of a person's attachment to their origin, and when their location of residence differs, that allows us to characterize it as a migration decision. We define the city-level out-migration rate as the ratio between those who leave their hukou city for more than six months, and the number of people whose hukou location is a given city.

Second, we construct an individual-level panel using the 2014 China Laborforce Dynamics Survey (CLDS), which records individual histories of location changes for a sample of 14,226 households across 29 provinces of China. The CLDS is a national longitudinal social survey, with information on education, work experience and migration. Since the survey asks retrospective migration histories of each individual, we are able to construct a longitudinal panel of location histories between 2000 and 2014. We define migration to be an indicator for whether an individual changed city locations between years, regardless of whether they change their *hukou* status. The CLDS allows us to account for individual-specific unobservables, track those who have moved multiple times and those who have moved and returned home.

We supplement the migration data with a measure of the stock of workers by skill level computed using the Census data. Migration choices ultimately affect the number of skilled and unskilled workers in each city. We will show that the ratio of skilled to unskilled workers in cities vary systematically with air quality. These changes in stock are the summary outcome of (net) migration decisions for all reasons and through all modalities (whether or not individuals change *hukou* status), and the object most sensible to use in our structural analysis for the quantification of productivity. Jointly, the three different migration measures we use either

[8] This definition is consistent with recent work on internal-migration in China (Combes et al., 2019; Tombe and Zhu, 2019). Only 7% of individuals have a *hukou* city that is not their birth city.

follows best practice, or improves on, the approaches to migration measurement in China implemented in the existing literature.

2.2 Air Quality Measures

We use satellite data to measure air quality, which has a few advantages over official sources of pollution data. First, satellite-based PM2.5 measures are available for all cities in China between 1998 and 2015, whereas official PM2.5 data are only available since 2012 for a small number of cities. Second, official air quality data may be subject to manipulation by local governments (Chen et al., 2012; Ghanem and Zhang, 2014). The satellite-based measure seems more reliable: We have compared it to monitor-based PM2.5 data collected by the U.S. Embassy and Consulates in China, and the correlation between the two measures is approximately 0.8.

City-level annual PM2.5 concentrations are measured using the Global Annual PM2.5 Grids derived from satellite data by Van Donkelaar et al. (2016).[9] They estimate ground-level PM2.5 by combining Aerosol Optical Depth (AOD) retrievals from the NASA MODIS, MISR, and SeaWiFS, which are subsequently calibrated to daily global ground-based observations of PM2.5 using Geographically Weighted Regressions (GWR). The raster grids of this ground calibrated PM2.5 data have a high grid cell resolution of 0.01 degree. Our data provide a comprehensive and reliable measurement of air quality for a wide range of cities in China, covering all the prefecture, sub-provincial, and provincial cities.

In robustness checks, we use the Air Quality Index (AQI) released recently by MEP to measure city-level air quality. Air Quality Index is an overall daily and hourly indicator for air pollution concentration calculated using multiple atmospheric pollutants including SO_2, NO_2, $PM10$, $PM2.5$, O_3 and CO. We calculate the annual mean AQI for each Chinese city based on the daily data.

[9]Fine particles ($diameter < 2.5\mu m$) are more hazardous than larger particles ($2.5\mu m < diameter < 10\mu m$) for mortality, cardiovascular and respiratory disease, and PM2.5 is considered to be the best indicator of the level of health risks from air pollution. For more background information see WHO report: http://www.who.int/mediacentre/news/releases/2014/air-quality/en/.

2.3 Inputs into Instrumental Variables

We obtain information on large-scaled power plants, their coal consumption, and plant-level electricity generation from China Electric Power Yearbooks and China Energy Statistical Yearbooks. Following Freeman et al. (2019), we designate thermal power plants as "large scale" if their installed-capacity exceeds 1 million KW. We combine these data with auxiliary information on the establishment year of these power plants, the angle between their locations and annual prevailing wind direction of each city, and the distance from their location to each city.

We collect data on thermal inversions from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), which records the 6-hour air temperature at different atmospheric layers. For each 6-hour period, we calculate the temperature change from the first to the second above ground atmospheric layer. If the temperature change is positive, a thermal inversion occurs and the difference in temperatures measures the strength of thermal inversions. We calculate the annual occurrence of thermal versions and the annual sum of thermal inversion strength using the 6-hour data.

Estimating the structural model requires us to develop a few other instruments. First, we aggregate firm-level data from the Annual Survey of Industrial Firms (ASIF)[10] to the city-industry level using the 2-digit industrial classification for manufacturing industries. This allows us to construct a measure of industrial composition of each city. Second, we derive information from a large-scale university expansion in China at the turn of the century that suddenly expanded college enrollment by 20% in certain cities to identify skilled-worker agglomeration effects. The data on the number of college students and college graduates at city level are from China City Statistical Yearbook. Third, we leverage variation in trade shocks to identify migration responses to wages. Data on Chinese trade are from the UC Davis Center of International Studies. The quantity and value of exports and imports by Harmonized System (HS) of product classification are available at the city-level. Data are available annually between 1997 and 2013, importantly covering periods before and after China's accession to the WTO in 2001. We construct city-level measures of baseline dependence on products more likely to be affected by tariff changes and trade policy.

[10]ASIF accounts for more than 90% of the total industrial output in China and over 71% of the industrial employment in 2004.

2.4 Wages, Controls and City-level Characteristics

Wage data are from the 2005 Census and the CLDS. Since the 2015 Census does not record individual-level wages, we use the CLDS to calculate city-and-education specific average wage.

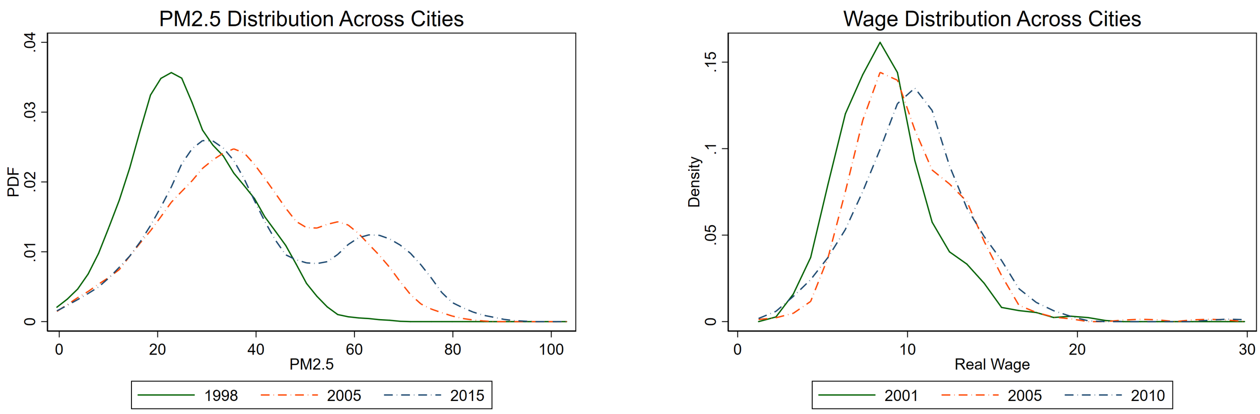
We collect city characteristics, such as population and GDP, from the China City Statistical Yearbooks. Weather data come from China Meteorological Data Service Center. We gather monthly data on average temperature, humidity, sunshine duration, and other weather amenities. We also calculate distances from each city to the three large seaports (Tianjin, Shanghai, and Shenzhen), and to the nearest big city and employ these variables as controls. Appendix Table B1 reports the summary statistics and a full description of the key variables used in the analysis.

3 Descriptive Patterns of Pollution and Migration

In this section we describe the spatial and temporal patterns of pollution, migration and wages in the raw data. These patterns motivate the hypotheses linking pollution variation to migration and wages, which we then subject to a more serious and rigorous inquiry in subsequent sections.

3.1 Spatial Dispersion in Pollution and in Wages

Figure 1: The Distribution in Pollution and Wages Across Cities



(a) Increasing Spread in Pollution Across Cities (b) Real Wage Distribution Across Cities

Notes: Distributions across cities for different years. Wage distribution across cities drawn from the China Statistical Yearbooks. Real wages are nominal wages deflated by local housing price. PM 2.5 data from the Global Annual PM 2.5 grids.

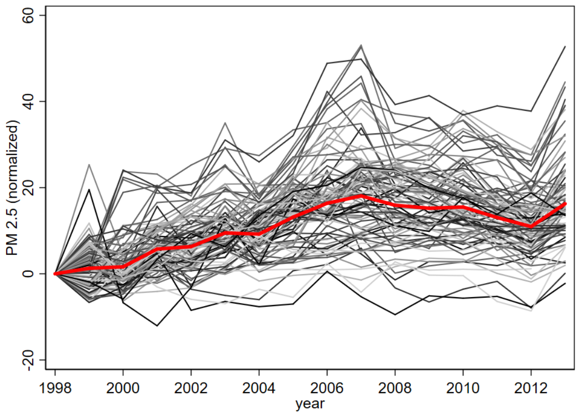
Figure 1 shows the increases in both the city-level spatial dispersion of PM 2.5 levels and in real wages over time. The left panel documents that pollution not only increased between 1998 and 2015, but it also became more variable across regions. The double-peak in the figure further indicates that the overall increase in PM 2.5 was driven by the emergence of some high polluting cities.

The right panel shows that both wages and the cross-city variance in real wages rise over time. If this implies an increase in the spatial dispersion of marginal products of labor, then that raises the possibility that moving workers from low marginal product cities to high marginal product cities may increase aggregate output.

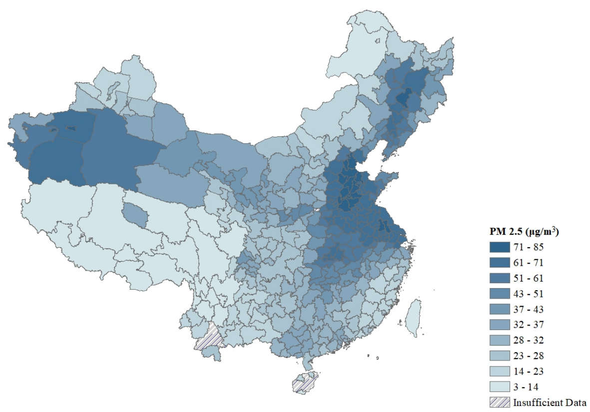
3.2 Spatial Distribution of Pollution and Migration Across China

Figure 2: The Distribution in Pollution Across Cities and Over Time

(a) Change in mean PM 2.5 over time



(b) PM 2.5 Concentration in China



Notes: The spatial and temporal distribution of PM 2.5 using the Global Annual PM 2.5 Grids. The map shows the geographic spread in 2015. Figure 2a shows the increase in PM 2.5 over time for the 100 largest cities in China, where the 1998 value of PM 2.5 is normalized to be 0.

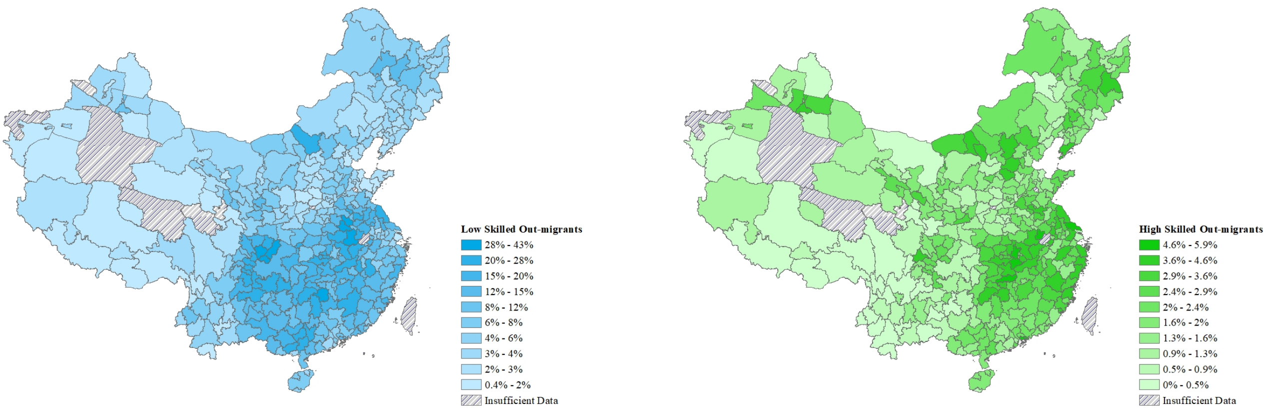
Figure 2a illustrates the time trend of annual PM2.5 concentrations in Chinese cities since 2001. The mean concentration exceeds WHO air quality guideline every year.[11] The figure also shows that the increase in the mean coincided with the increase in cross-city dispersion in pollution documented in Figure 1a. The increase in the overall mean was driven by dramatic increases in PM 2.5 in a subset of cities.

[11] See http://www.who.int/mediacentre/factsheets/fs313/en/ for more background information.

Next, we explore the spatial and temporal variation in air quality. Figure 2b illustrates the annual average satellite PM2.5 concentration data for 2015. Air quality is unevenly distributed across China. The coastal areas of North-east and eastern China experience the most severe air pollution. Manufacturing industries are concentrated in the east. The north-east further suffers from coal-burning due to heating needs, which exacerbates pollution even relative to high economic growth areas of the south.[12]

Figure 3: The Geographic Distribution of the Share of Out-Migrants by Skill

(a) Share of Low-Skill Out-Migrants (b) Share of High-Skill Out-Migrants



Notes: Low-skilled denotes people whose highest degree is high school or below. High-skilled denotes people whose highest degree is some college or above. Out-migrant shares are ratio of those who leave their hukou city for more than six months, and the number of people whose hukou location is a given city.

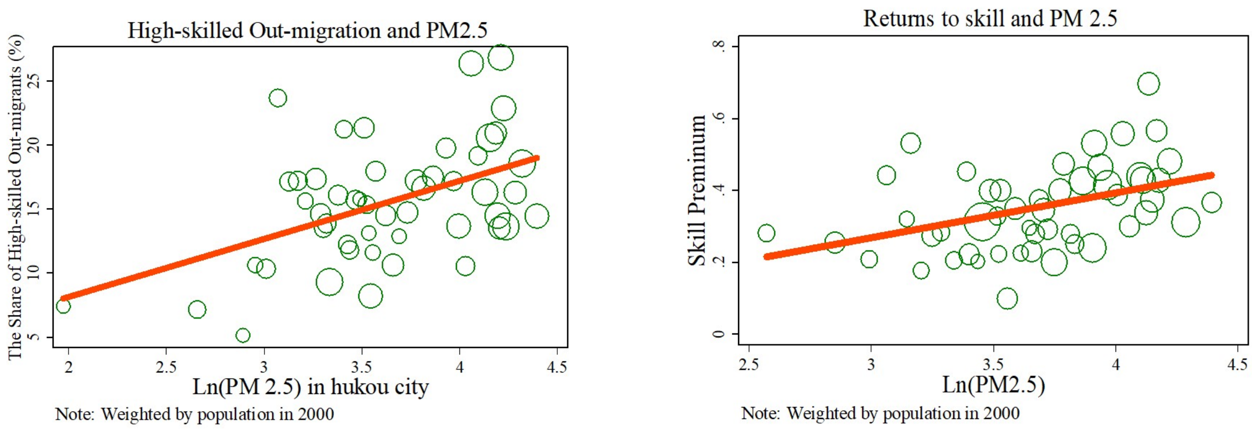
Next we examine the geographic patterns of emigration of low-skill (Figure 3a) and high-skill (Figure 3b) migrants. Low-skill emigration rates are very high in the south of China, while high-skill out-migrants are comparatively more populous in the north east and the east. Recall from Figure 2b that pollution is also relatively more concentrated in the north-east than in the south. These three figures therefore jointly indicate that pollution is more spatially correlated with high-skilled emigration rather than low-skilled.

Figure 4a explores whether that observed spatial correlation creates any city-level association between pollution and the share of emigrants who are high-skilled. There is a clear positive association, suggsting that the high-skilled are *relatively* more likely to leave polluted areas. We will explore this intriguing correlation more rigorously in subsequent sections, to identify

[12]Dust-storms in southern Xinjiang province are responsible for the isolated area of high particulate matter observed in the west. This area is otherwise not highly economically active.

whether the relationship is causal.

Figure 4: The Effects of High PM 2.5 at Origin Cities



(a) High-Skill Emigration Share and PM2.5

(b) Returns to Skill and PM2.5 at Origin

Notes: The share of high-skilled out migrants denotes the share of college (or above)-educated out-migrants from the city-level hukou population. Returns to Skill denotes the return to some college or above education.

Finally, Figure 4b examines the association between pollution and the wage returns to skills that emerges in each city. Returns to skill are higher in polluted cities. Economic theory provides a simple explanation for the two related correlations depicted in Figure 4: Higher out-migration of college workers in response to pollution makes the high-skilled relatively scarce in those cities, and in equilibrium, creates a compensating differential for poor air quality for skilled workers. This relationship will endogenously emerge in the general equilibrium model of pollution, migration and wages that we develop. Figure 4b also highlights a key insight about the benefits of pollution control policy that will emerge in our model: Reducing PM2.5 in highly polluted cities would induce high-skilled workers to move to the cities where their skills are relatively scarce, and this sorting could be a mechanism that raises aggregate productivity.

4 Identifying the Causal Effect of Pollution on Migration

Our main specification studies the effects of PM 2.5 concentration in city j on the amount of out-migration by skill group. Our primary regression of interest is as follows:

$$M_{ij} = \alpha + \beta Log(PM2.5)_j + \mathbf{X}\beta + \epsilon_{ij} \ , \tag{1}$$

where M_{ij} is an indicator for whether or not individual i left city j, and $\mathbf{X}$ are a vector of controls. This approach, however, may produced biased estimates of the causal relationship between PM 2.5 and migration decisions, as other city level unobservable characteristics may be associated with higher pollution levels and incentives to leave. Indeed, one may expect that pollution is strongly associated with the underlying structure of the economy, as polluted areas may have high manufacturing based economies. To get around these issues, we use a few different identification strategies to isolate the effect of pollution not driven by economic activity. In addition to OLS and panel fixed-effects models, we discuss in detail two different instrumental variables strategies. In Appendix A we include a set of robustness checks to help build confidence in our various strategies.

4.1 Instrument 1: Wind Direction and Coal-Fired Power Plants

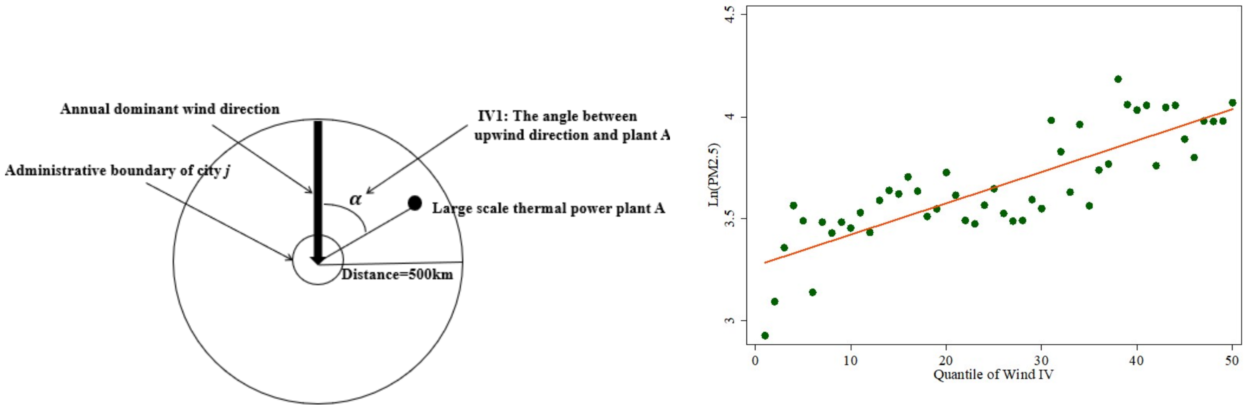
The first instrument we consider is based on recent work by Freeman, Liang, Song, and Timmins (2019), who use it to evaluate hedonic price models of willingness-to-pay for clean air in China. To formulate the instrument we use the inverse angle and distance weighted sum of coal consumption of distant large-scale coal-fired power plants. The underlying variation is driven by how wind patterns blow pollutants from distant coal plants to cities. Our first stage relationship is:

$$Log(PM2.5)_j = \gamma_0 + \gamma_1 \sum_{p}^{P} \left(\frac{1}{\alpha_p + 1}\right)\left(\frac{1}{d_{pj}}\right) C_p + \varepsilon_j \, , \tag{2}$$

where α_p denotes the angle between the annual prevailing wind direction of city j and the large-scale power plant p, d_{pj} is the distance between the plant p and city j, C_p is the annual coal consumption in plant p. We restrict our analysis to all thermal power plants that are located *outside a given city* within a 500km radius from the city center. Figure 5a explains the intuition behind the instrument. Our first-stage relationship in Figure 5b shows that cities downwind from, and closer to, higher coal-consumption power plants are more likely to be affected by poor air quality.

We expect that our instrumental variable is orthogonal to local economic activity. First, wind direction is naturally determined and as such, is unrelated to local economic attributes.

Figure 5: Wind direction, distance, and production in thermal power plants



(a) Depiction of instrument (b) First-stage relationship

Second, the large-scale thermal power plants supply electricity to vast areas of China; some of them do not even supply electricity to their nearby cities, but rather to many remote provinces. Third, in China, the allocation of electricity supply from large-scale power plants is determined by the central government. Although many reforms have taken place over the past 30 years, there are still strict regulations in the power sector and ownership of the sector is largely with the state. The central government owns the grid, and controls the setup and operation of power plants if their generating capacity is large. Thus, local governments find it difficult to exert influence on the setup of large-scale power plants and the allocation of electricity supply from them. Finally, the impact of distant power plants on local economic activity is extremely small, but the particulate matter spewed from coal-fired power plants located at upwind region contribute substantially to local air pollution.

We examine threats to using this instrument as identifying variation in Appendix A.1.2. We consider whether the location of power plants may depend on the simultaneous combination of wind direction, distance to the plant, and the amount of coal consumed. For instance, if we are concerned that newly built plants are are placed away from important cities, we show robustness to excluding power plants in a 200km radius away from cities, robustness to using only old power plants, and excluding richer or capital cities, among other specification tests. We also conduct numerous falsification tests showing that baseline city characteristics do not predict the future placement of plants, and plants that are upwind or orthogonal to the wind direction do not affect air quality nor migration.

4.2 Instrumental Variable 2: Thermal Inversions

Our second instrument uses the number and strength of thermal inversions, which has in the past been used to predict air quality in Mexico (Arceo et al., 2016), the US (Hicks et al., 2015) and Sweden (Jans et al., 2014), among other settings. Most recently, Chen et al. (2017) show that the number of thermal inversions predicts the movement of people across China as well. We build upon their work which shows that those with higher levels of education are more responsive to poorer air quality, by using newer migration data from the 2015 Census at the individual level (rather than quantifying migration from population changes).

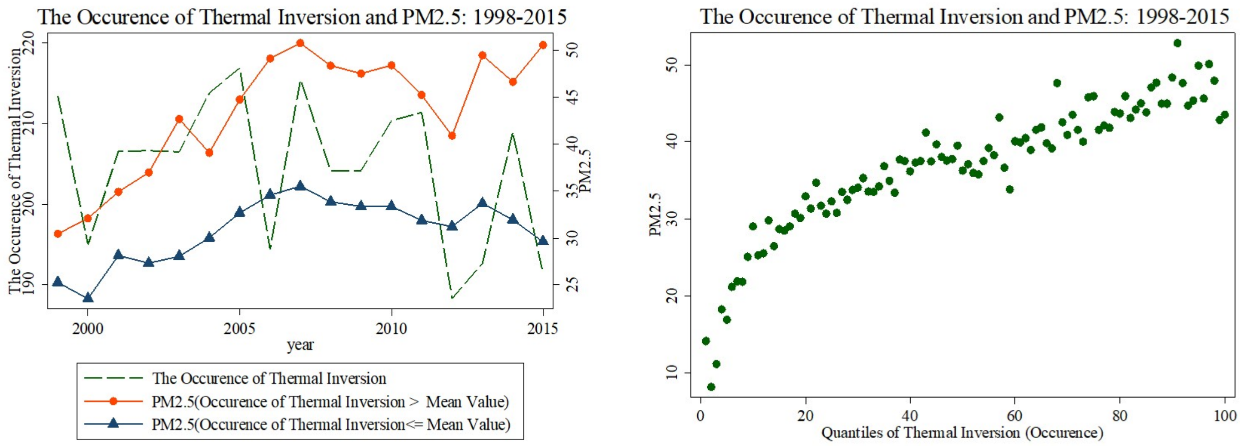
A thermal inversion is the difference between the above-ground temperature and the ground temperature, which traps pollutants, and is a strong predictor of poor air quality. We create two measures of such inversions TI_d in city d. First, we use the decennial Census across two decades (1995-2015), and count the number of thermal inversions in each decade t. Next, we measure the strength of these inversions and create a measure of the mean strength over the decade. We use this measure in both the cross-section, and in panel form. Our panel specifications include decade fixed effects τ_t and city fixed effects δ_d:

$$Log(PM2.5)_{dt} = \gamma_0 + \gamma_1 TI_{dt} + \delta_d + \tau_t + \varepsilon_{dt}$$

As polluting potential rose over time in China, areas with more thermal inversions trap the pollutants in the nearby atmosphere. This can be seen in Figure 6, which shows that there is a lack of any trend over time in thermal inversions, yet as the cities pollute more over time, areas with more inversions see a sharper rise in poor air quality. The right side of the figure shows the strong correlation between inversions and PM 2.5 suggesting that this relationship has a strong first stage.

Our preferred specification leverages the panel variation. We aggregate inversions over the preceding five years to study the impacts on out-migration, as migration decisions may take time to manifest. The advantage of the panel specification is that we can show numerous robustness checks, such as how lagged pollution levels do not predict future inversions. Together, these two instruments capture the variation in air quality due to either wind direction or meteorological

Figure 6: Thermal Inversions and Air Quality



phenomenon, and are less likely to be directly related to production.

5 Empirical Results

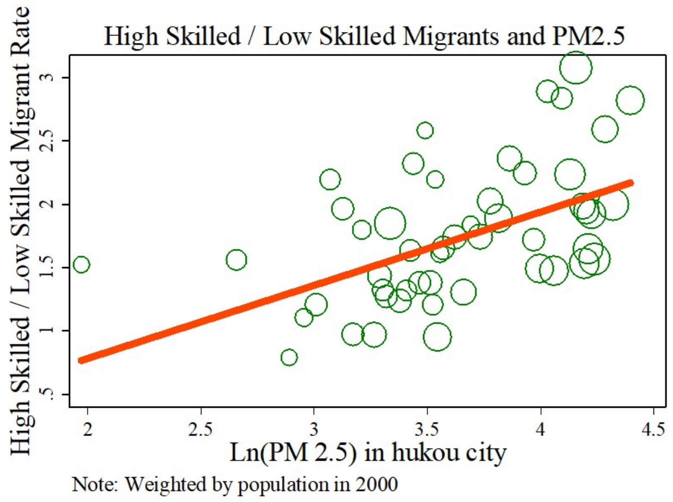
Here, we discuss our main empirical relationships between migration and pollution. We relegate most of the extensive discussion on our empirical results to Appendix A, which includes a long set of specification tests, falsification tests, sample restrictions, different panel-data structures, and different types of controls.

5.1 The Relationship Between Pollution and Migration

Figure 7 captures the relationship between differential mobility and pollution in the raw data. In the vertical axis we plot the ratio between the out-migration rates for college educated workers and the rates for non-college workers. This ratio increases with higher PM 2.5 levels in source cities, suggesting that the out-migration for college workers is a lot stronger than that of non-college workers.

In Table 1 we examine the OLS relationship between PM 2.5 and out-migration, measured with the help of an individual-level indicator of leaving ones *hukou* city. Everywhere, we divide the sample into those with a college degree and those without. Our estimates suggest that a 10% increase in PM 2.5 raises out-migration rates by 0.497 percentage points, with the effect being meaningfully larger for those with a college degree (0.655 percentage points) that those

Figure 7: The Ratio of High-to-Low Skilled Out-Migration



Notes: High/low skilled out-migrants denotes the ratio of high-skilled (some college or above) out-migrants to low-skilled out-migrants (high school or below). We plot these ratios in 2015 against PM 2.5 levels in 2015.

without (0.452 percentage points).

Table 1: Pollution and Out-Migration (OLS)

	Dependent variable: Leave hukou city indicator		
	Full sample	Low edu	High edu
Log(PM 2.5)	0.0497***	0.0452***	0.0655***
	(0.0121)	(0.0128)	(0.0119)
Observations	462,046	384,221	77,825
Adjusted R-squared	0.030	0.029	0.080
City Controls	Y	Y	Y
Demographics	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

In Appendix A.1.1 we study our results across a few different specifications, including the instrumental variable models. In Appendix Table A1 we show the first stage of the different instrumental variables, all of which display a strong and robust relationship between our instruments and atmospheric PM 2.5. The 2SLS versions for which appear in Tables A2 - A4. The results once again show that there is more out-migration among higher educated workers than those with less education. In Table A4 we try combinations of the different instruments.

Tables A5-A6 show that the in-migration decisions also depend on destination PM 2.5.

Together these results suggest that workers leave more polluted areas and seek out less polluted cities. Importantly, this response to levels of pollution is stronger among higher educated workers than it is among the less educated. Fewer skilled workers will tend to raise the skilled wage and lower the unskilled wage. As a result, moving a skilled worker from a city that has a lower marginal product of skill to a city with a higher marginal product, will lead to an increase in aggregate output.

5.2 Robustness and Heterogeneity

In Appendix A.1.2 we examine threats to identification for our various sources of variation. We seriously consider the claim that the location of power plants may depend on the simultaneous combination of wind direction, distance to the plant, and the amount of coal consumed. First, we may think that for power plants placed close to big cities, policy makers may rely on wind direction and access to coal, to make decisions on where to place them. Indeed, they may even choose to place them far away. In Table A7 we create an instrument that does not take into account plants placed either in a 200km or a 400km radius, and find that, if anything, our results are more precisely estimates.

We may also be concerned that newly built plants are placed in areas based on wind and distance models. So, in Table A8 we only restrict our sample to old plants, and find similar effects. In Table A9 we show that baseline population, GDP and electricity consumption do not predict future upwind plants, or future iterations of our IV. Suggesting that it is not that they avoid richer, influential cities when building plants, and that plants are not built in areas that have a higher need for electricity at baseline – in any case, most electricity is directly supplied to the larger grid.

Similarly, for our thermal inversion instruments, we show that lagged pollution levels do not predict future inversions. Indeed, even lagged inversions do not predict future ones – suggesting that their occurrences are hard to predict.

In Appendix A.1.3 we explore different panel data structures for our primary relationship of interest. Panel data structures allow us to use fixed effects that control for unobservable time-invariant heterogeneity, and may provide more robust estimates of the migration response to pollution. First, we show that pollution decisions may react to changes in long-term pollution averages in Table A11. Then we move from individual level data, to a city-level panel to run fixed-effects regressions. In doing so, we need to explore the time structure of the data a little better. In Table A12 we show that contemporaneous changes in pollution actually do not seem to affect migration decisions. This may reflect the fact that since migration is an important long term decision, immediate shocks may not induce people to move. In contrast, in Table A13, we measure PM 2.5 over the 5 years preceding a period to capture migration decisions in the face of persistent increases in pollution. In Tables A14-A15 we use the thermal inversions instrumental variables and examine this relationship over a longer time period. Since our decision to use 5-year averages is arbitrary, in Tables A16-A18 we explore other aggregations over time.

We summarize our analysis of the different sources of variation in Appendix 5.3, where we show that consistently across specifications, there is an increase in out-migration among the high skilled, but no corresponding increase among the low skilled.

In Appendix A.1.4 we explore the variation in pollution driven by China's Huai river heating policy. As Chen et al. (2013) show, heating policy in China generated an artificial discontinuity in air quality on two sides of the Huai river. North of the Huai river, the government established free winter heating of homes, and provided free coal and boilers to residents. Even in the 2000s, there is a sharp discontinuity in the use of boilers for heating, leading to a discontinuity in air quality across the Huai river. In Appendix A.1.4 we examine the consequences of this policy on migration. While we fail to find differential out-migration in response to pollution, we do find stark differences in in-migration between skilled and unskilled workers in response to air quality differences.

We then turn our attention to studying different model specifications, sub-samples, and checking the robustness of our estimates to different controls. Appendix examines our main relationship across different sub-samples of the data. First, in Table A19 we use a different source of data, and a different definition of migration to replicate our results. We use the China Labor Dynamic Survey (CLDS) that asks individuals for a retrospective history of location changes, allowing us to construct an individual panel of migration decisions irrespective of changes in *hukou* locations. Table A19 shows that our results are similar to our analysis using the Census data.[13]

Next, we explore different ways to disaggregate the data. In Table A20 we disaggregate the education levels into more categories and see a sharp education gradient in out-migration responses: those with more education are more responsive. Table A21 studies heterogeneity across rural-urban status, while Table A23 shows that the youth are more responsive to pollution.

Two other important tables address any lingering endogeneity concerns with the wind direction and coal-fired power plants IV. In Table A22 we exclude large, influential cities, cities that pollute a lot, and major province capitals, to account for any differences in political influence or outliers in the access to skilled jobs. In Table A24 we drop coal-producing regions so as to allay any concerns that plants are more likely to be located in such regions, which in turn have different economies.

Appendix studies other model specifications underlying the wind direction and coal plants IV. First, in Table A25 we decompose the IV, and exclude components such as the distance to the nearest plant, and the amount of coal consumed, to isolate variation coming from the other components of the interaction term. Then, in Table A26 we restrict our sample to only upwind plants – this specification basically compares similar cities but one with an upwind plant and another with a plant downwind. The last few columns explicitly models the IV to be the ratio of upwind to downwind plants, to solely rest on the variation from upwind plants. In Table A27 we extend this analysis by creating various 'placebo' instruments based on artificially changing the wind direction and showing that these falsified instruments do not predict pollution levels, nor in-migration.

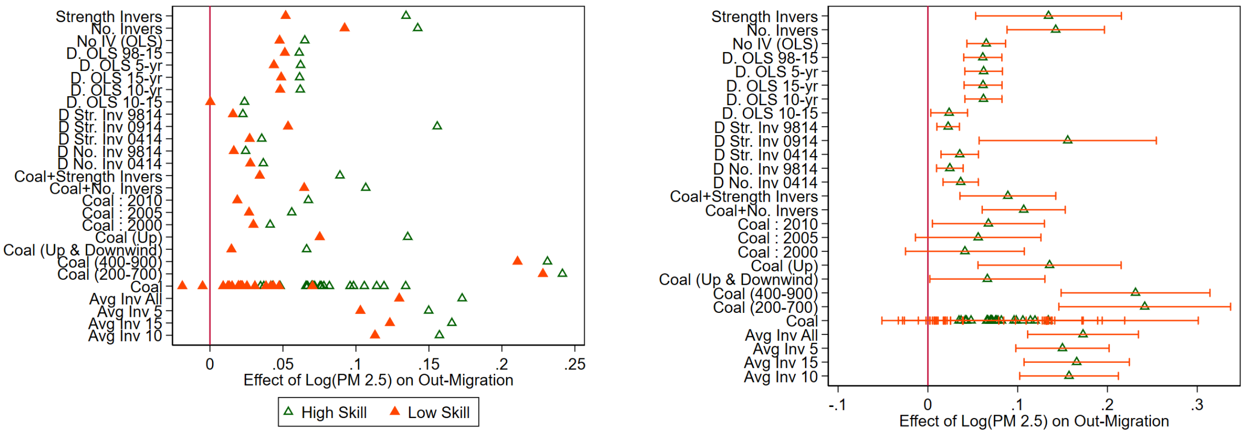
Finally, Table A28 uses the Air Quality Index (AQI) as the endogenous variable of interest. Since PM 2.5 may be correlated with other pollutants, these instruments affect overall air quality, and we may be picking up the combined impact of many pollutants. We show that our results are robust to using AQI as our independent variable of interest. Appendix A.2.3

[13]The CLDS data also show that among migrants, high-skill workers are far more likely to change *hukou* locations, as expected. This suggests that our Census definition of migrants as those with different *hukou* locations may produce conservative estimates of the overall location changes of high-skill workers. Finally, when we estimate our model, we only rely on how the skill-biased stock of workers vary with pollution, as it encapsulates net migration flows and counts all kinds of migrants.

highlights robustness to a long list of controls, including electricity consumption and baseline skill-distribution indicators (Table A29), baseline economic indicators, natural amenities and weather (Table A30). These controls do not qualitatively affect the main patterns we observe. In Figure 9 we summarize all the different model specifications into two figures that consistently show the effects of pollution on skilled out-migration.

5.3 Summary of Alternative Specifications

Figure 8: Different Sources of Variation



Summary of results using different sources of variation. We compile coefficients from different specifications. On the left we show both the coefficients on high and low skill workers. On the right, we concentrate on college educated workers, and include 90% confidence intervals. The "Coal" IV specification was run many times with different sets of controls (discussed in the following section), and so has many point estimates.

Across different specifications, a simple pattern emerges: in response to poor air quality, there are meaningfully responses to high-skill out-migration, but smaller impacts on low-skill out-migration. We summarize these results in Figure 8 which show not just the many specifications discussed above and others in the appendices, but also additional specifications for which we do not show tables.

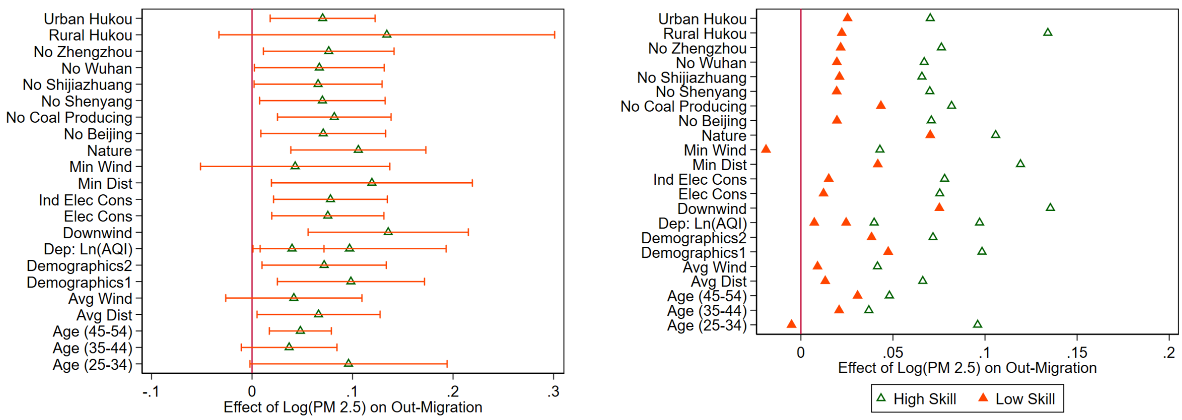
The specification that relies on wind direction and coal-fired power plants ("Coal" for short, in the figures) was subject to an additional set of tests discussed in the next section, and so has many point estimates under display. The first two specifications for the strength and number of inversions are simply IV results discussed above. Then we show the OLS specification, before going into various panel-data fixed effects models (each with "D." label). After the panel models, we show results for the combined coal and inversions instruments.

The next set of results support our causal claim. Here, we focus on the coal-fired power plants instrument, and limit ourselves to plants built up to a certain point of time. For instance, “Coal: 2000” relies on plants built before 2000, and excludes newly built power plants. We do this so as to allay any concerns that newly built plants may be placed endogenously simultaneously based on wind directions, distance to cities and access to coal.

The “Coal (Up)” results only restrict the IV to rely to directly upwind plants. The “Coal (400-900)” IV excludes any plants built within a 400km radius from a city and instead only captures plants built 400-900 kms away. This is done to allay concerns related to the endogenous placement of plants in close proximity to cities. Finally, the “Avg Inv” specifications create IVs based on the the average number of inversions over a 5, 10 or 15 year period.

Together these results show two sets of patters: first, the effects on high-skill workers are always larger than low-skill workers, and second, the effects on high skill workers are always positive (i.e. more out-migration) and often statistically different from zero.

Figure 9: Different Samples, Controls and Models



Summary of results using different samples, controls and models. We compile coefficients from different specifications. On the right we show both the coefficients on high and low skill workers. On the left, we concentrate on college educated workers, and include 90% confidence intervals.

In Figure 9 we summarize our results for alternative samples, controls and model specifications, with the help of some figures. The results from these figures are based on the tables in the appendices, along with some additional results. In addition, it includes robustness to specifications that use the minimum distance and minimum wind direction angle when constructing the coal-fired plants IV (instead of the averages). Consistently, the left hand side graph shows that the effect on high skill worker out-migration is strong and statistically different from zero. The right side figure shows how the effects on high skill worker migration is always consistently larger than that for low skill out-migration.

5.4 Wage Returns and Pollution

If sorting based on skill levels leads to a geographic re-allocation of skill, we should expect that the returns to skill differ across regions. Skilled workers are going to be scarce in regions that they leave, raising their value on the labor market in such regions. Additionally, given the complementarity between skilled and unksilled workers, regions that lose skilled workers will have less productive unskilled workers. As such, regions that lose skilled workers would have higher skilled wages, lower unskilled wages, and therefore higher returns to skill.

In Table 2, we examine this possibility empirically. We estimate the simple Mincerian returns to skill in each city. Then we create a measure that is the difference in returns between destination and hukou cities, and the differences in PM 2.5 between destination and hukou cities. We show the IV-2SLS relationship between these variables in Table 2. For the columns where we look at secondary education, we use the returns to secondary education as an outcome, and exclude the sample of those with a college degree. When studying the returns to college education, we exclude those with only a primary level of schooling.

Table 2 displays a simple fact: wage returns are higher in regions that have more pollution. This is consistent with the fact that differential out-migration of skilled workers raises the relative marginal product of skilled (to unskilled) work. We formalize this result in our theoretical framework below. Indeed, we find that the returns to skill are substantially higher in polluted regions (compared to less polluted cities).

Table 2: Difference in Wage Returns by Differences in Pollution Levels

	Dependent variable: (Education return in residential city) - (Returns in hukou city)			
	Secondary education	Tertiary education	Secondary education	Tertiary education
	Full sample		Migrant Sample	
Δ ln (PM2.5)	0.131**	0.554***	0.133**	0.558***
	(0.0616)	(0.0916)	(0.0611)	(0.0929)
City Controls	Y	Y	Y	Y
Demographics	Y	Y	Y	Y
N	75397	69310	8251	7404
R2	0.556	0.000	0.437	0.000

Notes: Independent variable of interest is the different in pollution rates between a pair of cities. Primary education denotes workers whose highest degree is below high school. Secondary education denotes workers whose highest degree is high school. Tertiary education denotes workers whose highest degree is some college or above. Controls are at the city (like distance to seaports) and individual (like gender, age) level. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

5.5 Hukou Restrictions, Preferences and Costs by Skill Level

We aim to understand what drives the differences in out-migration rates by skill. China's *hukou* restrictions make it easier for skilled workers to move to certain cities and have access to jobs and public services. In Table B2 we show a few examples of such restrictions, highlighting how education levels can help one gain enough *hukou* points to be eligible. Such restrictions make it costly for unskilled workers to move, and are possibly a contributing factor in driving these differential migration rates. Indeed, as *hukou* restrictions make it difficult for the unskilled to leave with the skilled, they may create an artificial mismatch between potentially complementary workers, and exacerbate the productivity losses due to worker re-allocation. Secondly, the costs of mitigating the impacts of poor air quality may also differ by skill level, as richer households may be able to afford air filters.

Last, preferences may play an important role. Pollution may matter less for unskilled workers as they may be trying to make ends meet with their lower wages. On the other hand, high skill workers may be more affected by city level amenities and as such, respond to changes in amenity values, such as pollution levels. In Table B3 we use the China General Social Survey (CGSS) which asks questions about whether the respondent thinks that the environmental issue in China is "terrible" or not. Here, the omitted category is those with less than a secondary education. We find that those with more education are more likely to claim that the environmental issue in China is "terrible." In Table B4, we use the same survey to explore not just the concerns for environmental issues, but also the actions taken on environmental issues. Once again, across the board, individuals with more education are more likely to discuss environmental issues, make donations for environmental protection, and make appeals or raise concerns on environmental problems. These differences at the skill level are statistically distinguishable from each other. Those with tertiary levels of education are meaningfully more concerned about environmental issues than those with primary education, and also than those with secondary education. Together, these results imply that the differences in migration patterns in response to changes in pollution partly reflect the differences in preferences for environmental quality, and partly the *hukou* restrictions.

6 Theoretical Framework

We use a simple theoretical framework to aid our quantification of the productivity consequences of pollution-induced migration. The model captures a few key features necessary for this quantification. First, it endogenizes the compensating differential, as experiencing pollution lowers the utility of workers in a manner that differs by skill. Second, *hukou* restrictions make it costly for workers to move to certain cities, and these costs also vary by skill level. Together, these contribute to the empirical patterns that show a differential out-migration by skill level.

Third, as college educated workers leave polluted cities, the marginal product of skilled labor rises. If skilled and unskilled workers are complements, with an elasticity of substitution σ_E, the marginal product of unskilled work falls. This leads to differences in the skill-wage premium, consistent with the empirical result that the returns to skill are higher in regions that have more pollution.[14] Moving a skilled worker from a low wage city to a high wage city will increase aggregate productivity.

An additional source of geographic-specific returns to skill is driven by the fact that some cities have more skill-biased capital, as captured by a city d specific and skill s specific productivity term θ_{sd}. Furthermore, the changing structure of skills in a city affect production and pollution levels. Skilled workers may induce either more or less pollution-intensive industries

[14]Even though unskilled workers may wish to follow skilled workers and leave, *hukou* restrictions may make it costly for them to do so.

to expand, and as such change the quality of air in the city. This feedback effect of migration patterns on where pollution takes place affects subsequent migration, which in turn affects production, and so on. Finally, agglomeration forces may increase aggregate productivity if skilled workers converge to high amenity cities, but house prices may also respond to such movements creating congestion in such cities.

Our framework generates simple estimable equations that we identify using instrumental variables. The main results will rest on a few different elasticities that allow us to perform a quantitative counterfactual where a reduction in pollution re-allocates skilled work to where the returns are higher. We allow for direct productivity effects of pollution which affect all workers, but we assume that the effect of pollution on productivity is not skill-biased.

6.1 Production and Labor Demand

Aggregate output Y_d in destination city d depends on L_d (effective labor), K_d (capital), and A_d (TFP). TFP varies across cities and with the lack of pollution Z_d.[15] Capital is perfectly elastically supplied across cities at rental rate R^*.[16] Effective labor supply L_d depends on labor L_{sd} at each skill level s.

$$Y_d = A_d L_d^{\varrho} K_d^{(1-\varrho)} \quad where \quad L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E - 1}} \quad and \quad A_d = \bar{A}_d Z_d^{\phi} \tag{3}$$

$0 < \varrho < 1$ is the share of output accruing to labor, $\theta_{sd} > 0$ is the productivity of workers with skill level s, and $\sigma_E > 0$ is the elasticity of substitution across skill groups. $\bar{A}_d$ is exogenous city-level productivity (fertile soil, rivers, land etc.), including the amount of industry-specific innovation. ϕ determines the elasticity of pollution with TFP. If $\phi < 0$, then pollution is associated with more productivity. If $\phi > 0$, pollution lowers the productivity of all workers.

The skill-biased productivity parameter θ_{sd} captures the productivity of each skill level, and increases with an increase in skill-biased capital in the city k_{sd}, such that $\theta'_{sd}(k_{sd}) > 0$.[17] Notice, it could also capture the fact that some cities have policies that raise wages for skilled

[15] Keeping with convention in the literature on amenities, higher Z_d means more amenities, so less pollution.
[16] The perfectly elastic capital assumption is not essential and can be relaxed.
[17] For completeness, one can explicitly model skill-biased capital within the nested CES framework and show how flexible ways of incorporating it do not affect the estimation or results, but there is no way to separately identify the elasticity of skill-biased capital from the direct productivity contribution of skill.

workers, or other relative labor demand shocks. The value of θ_{sd} therefore varies across cities because of the variation in skill-biased capital k_{sd}, and other factors that make skilled work more productive in d. The average log earnings for skill s in destination d are:[18]

$$\log w_{sd} = \log\left(\frac{\partial Y_d}{\partial L_{sd}}\right) = \log \widetilde{\varrho} + \log \theta_{sd} + \frac{1}{\sigma_E}\log Y_d - \frac{1}{\sigma_E}\log L_{sd}\ , \tag{4}$$

where $\log \widetilde{\varrho} \equiv \left[\left(1-\frac{1}{\sigma_E}\right)\left(\frac{1-\varrho}{\varrho}\right)\log\left(\frac{1-\varrho}{R^*}\right)\right]$ is common across all cities and workers.[19] There are a few components that drive the differences in average earnings when comparing two different skill-groups s in two different labor markets d represented in Equation 5:

$$\log\left(\frac{w_{sd}}{w_{s'd'}}\right) = \underbrace{\log\left(\frac{\theta_{sd}}{\theta_{s'd'}}\right)}_{\text{productivity}} + \underbrace{\frac{1}{\sigma_E}\log\left(\frac{Y_d}{Y_{d'}}\right)}_{\text{output}} - \underbrace{\frac{1}{\sigma_E}\log\frac{L_{sd}}{L_{s'd'}}}_{\text{skill distribution}} \tag{5}$$

This equation captures why firm-level demand may drive systematic differences across people and across labor markets. The first component – ‘productivity’ – θ_{sd} is the higher productivity associated with more education. Not only are skilled workers more productive, but variation in the supply of skill-biased capital across cities affect earnings. The second – ‘output’ – is the difference across labor markets related to differences in the size of the economy, including TFP. The third – ‘skill-distribution’ – is the difference in earnings due to differences in the supply of more educated workers L_{sd}.

Equation 5 is the relative labor demand curve and highlights the importance of elasticities: how much the skill distribution affects the difference in earnings depends on the elasticities of substitution σ_E. In regions that have relatively more skilled workers, the skill premium will be relatively lower. Whereas for regions with more skill-biased capital, the skill premium is higher. Finally, skill-biased migration will change these quantities and affect skill-premia.

Additionally, we assume that economic production produces pollution as well. As the skill mix changes across cities, some cities produce more output than others, which in turn raises the amount of pollution produced. To be specific, the increase in pollution depends on two

[18]This is at the optimal value of K_d^*, so that $Y_d = A_d\left(\frac{1-\varrho}{R^*}\right)^{\frac{1-\varrho}{\varrho}} L_d$.

[19]For tractability, we have ignored the role played by changes in prices. It is straightforward to include a $\log P_d$ that will be associated with the $\frac{1}{\sigma_E}\log Y_d$ term, and not affect relative skilled-unskilled wages.

things. First, the aggregate output, which in turn captures the size of the economy, industrial production and congestion. Second, the skill-mix, which captures the differences in types of production (industry vs services), changes to local laws, and amenities produced as the skill mix changes.

$$Z_d = Z(Y_d, L_{sd}, L_{ud}, \bar{Z}_d) = \bar{Z}_d \left(\frac{L_{sd}}{L_{ud}}\right)^{\psi_1} Y_d^{\psi_2} \tag{6}$$

6.2 Migration and Labor Supply

We assume that workers have preferences over locations, either because of tastes (some prefer to be closer to home, while others prefer big cities), or because it is more 'costly' for some people to migrate and leave home. The indirect utility of worker j, with skill group s, in destination d, and migrant status k is given by:

$$V_{sjdk} = \epsilon_{jd} w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d exp^{-H_{sd}} \ , \tag{7}$$

where ϵ_{jd} is a random variable measuring preferences for a specific city d by individual j. A larger ϵ_{jd} means the worker is particularly attached to a city d. Migrant status k is either m for migrants and n for natives. Similarly, H_{sd} captures the *hukou* costs of migrating to the city for migrants. As such $H_{sd} = 0$ for natives. hp_d are housing prices at the city level, and ν_s capture the share of expenditures on housing by skill-level.

The important addition here is Z_d, which is a measure of the lack of pollution. a_d represents other non-pollution related amenities. The compensating differential elasticity will be captured by γ_s, and varies by skill level.[20] Here, marginal workers are those that are indifferent across cities, and as such are likely to be induced into migration by pollution. Inframarginal workers have higher utility in the city they live in currently, than in all the other cities.

We assume that ϵ_{jd} are independently distributed and drawn from a multivariate extreme

[20] If $\gamma_s > \gamma_u$, the lack of pollution is a normal good.

value distribution (Eaton and Kortum, 2002). The joint distribution of ϵ_{jd} is given by:

$$F(\epsilon_1, ...\epsilon_D) = exp(-\sum_{d}^{D} \epsilon_d^{-\eta}) \ , \tag{8}$$

where $\frac{1}{\eta}$ determines how strong the idiosyncratic location preferences are, and so how responsive workers are to wage or pollution changes. If location preferences are very strong, then workers may not migrate even when wages differ widely, or pollution levels are high.

Mobility determines that workers move to places where their utility is higher, implying that given costs of moving, there are no arbitrage opportunities available. Local ties and *hukou* migration costs are captured by ϵ_{jd} and H_{sd} respectively. A person chooses city d over d' if $V_{sjdk} > V_{sjd'k}$. In Appendix C we derive an expression for labor supply and average city-utility. The number of skilled workers in city d is given by:

$$L_{sd} = \frac{\left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d\right]^{\eta}}{\overline{V_{sd}}^{\eta}} \quad , \tag{9}$$

where $\overline{V_{sd}}$ captures the average utility of being in city d. Note that because of migration costs, utilities are not equalized across cities, and as such the term has a d subscript. For instance, if a high-amenity city has a very restrictive *hukou* policy it may have a high average utility as not enough people can enter and lower wages and raise house prices. Yet, higher *hukou* restrictions will lower the utility for all individuals in other cities, as their option value of moving to a potentially desirable location falls. Therefore, as we show in Appendix C that $\overline{V_{sd}}(H_{sd})$ is directly a function of city-specific *hukou* restrictions. Taking logs of Equation 9:

$$log\ w_{sd} = log\overline{V_{sd}}(H_{sd}) + \frac{1}{\eta} log L_{sd} - \gamma_s log\ Z_d + \nu_s log\ hp_d - log\ a_d \tag{10}$$

It is now quite clear that $\frac{1}{\eta}$ is the elasticity of labor supply. If workers are attached to their location, or migration costs are high, then workers will not move even if pollution is high or wages are low. Finally, pollution Z_d at the city level can be separated into pollution that depend endogenously on production and skill-mix Z_d, and a part that depends on other plausibly exogenous factors $\bar{Z}_d$, such as thermal inversions and being down-wind from power plants, as in Equation 6.

6.3 Agglomeration, House Prices and Hukou Costs

In an extension, we consider the role played by agglomeration economies. We expect the share of skilled workers to raise TFP levels in the city via non-excludable innovation (Arrow, 1962).

$$A_d = \bar{A}_d Z_d^{\phi_1} \frac{L_{sd}}{L_{ud}}^{\phi_2} \tag{11}$$

Furthermore, like in Moretti (2011) we assume a simple housing market of the form $hp_d = N_d^{\psi_3} \frac{L_{sd}}{L_{ud}}^{\psi_4}$. More people in the city raise house prices, and a richer population leads to higher prices as well.

Hukou costs affect the welfare of both migrants and non-migrants, and as such shift the labor supply curve. To empirically understand the role played by *hukou* costs, we create a *hukou* index that depends on each city's *hukou* policy, and parameterize the term in the following manner, where ζ_s contains other skill-level characteristics that enter the equation for $\overline{V_{sd}}(H_{sd})$, as shown in Appendix C:

$$log\frac{\overline{V_{sd}}(H_{sd})}{\overline{V_{ud}}(H_{ud})} = -\lambda_1(Hukou_{sd} - Hukou_{ud}) + Log\ \frac{\zeta_s}{\zeta_u} \tag{12}$$

6.4 Equilibrium and Elasticities

Equations 3- 12 charachterize the model's equilibrium, which can be described as a set of wages, amenities, house prices, migration costs and labor allocations, such that workers are paid their marginal product, and workers choose cities to reside in.

In equilibrium, the labor market clears for each skill level $\{s, u\}$. The supply of L_{kd} equals the demand for L_{kd} for all d, and all $k = \{s, u\}$. Income produced in a city is consumed in the city d, and there are no savings. Aggregate output is simply the sum of output in each region $Y = \Sigma_d^D Y_d$.

When there is a fall in pollution, there is in-migration of skilled workers (or less out-migration of skilled workers). This shifts out the relative labor supply curve, such that the difference in wages is the compensating variation (in partial equilibrium). In (general) equilibrium though, skill-premia are different across the two regions because of migration, and will play an crucial role in the quantification exercise. A few important important elasticities help determine the size of theses effects, such as γ_s (the compensating differential elasticity), η (the migration-cost / preferences elasticity, or labor supply elasticity), σ_E (the elasticity of substitution across skill-levels in production, or the relative labor demand elasticity) and ψ_1 (the pollution response to changing skill-mix).

These elasticities are have meaningful implications for aggregate output, and help determine the productivity consequences of pollution in conjunction with the other elasticities.

7 Estimation of Model Parameters

To measure the productivity and welfare consequences of pollution-driven skill-biased sorting, we need to estimate the following city-level parameters: $\{\theta_{sd}\}$ and the following aggregate elasticities: $\{\eta, \sigma_E, \gamma, \psi_1, \psi_2, \psi_3, \psi_4, \lambda, \phi_1, \phi_2\}$. We estimate these parameters based on city-level relationships for a smaller set of largest cities which have consistent data on all the required variables across years. In our preferred estimates we weight by baseline population in 2000 and control for the city characteristics as in our earlier results, but show robustness to alternative specifications.

To aid estimation, we derive a few different versions of our model equations. Since σ_E helps determine the change in relative skill-unskill wages in response to a change in relative skill-unskill workers, we can derive a relative demand curve where within a city, the size of a city's output (and other city-level characteristics) are differenced out, as in Equation 13:

$$log \frac{w_{sd}}{w_{ud}} = log \frac{\theta_{sd}}{\theta_{ud}} - \frac{1}{\sigma_E} log \frac{L_{sd}}{L_{ud}} \tag{13}$$

Similarly, we can reformulate the labor supply, as in Equation 14:

$$log \frac{w_{sd}}{w_{ud}} = log \frac{\overline{V_{sd}}(H_{sd})}{\overline{V_{ud}}(H_{ud})} + \frac{1}{\eta} log \frac{L_{sd}}{L_{ud}} - \gamma log\, Z_d + \nu log\, hp_d, \tag{14}$$

where $\gamma = \gamma_s - \gamma_u$ and $\nu = \nu_s - \nu_u$. As such, we will not estimate γ_s and γ_u separately, but rather just the difference between the two.

The labor market equilibrium (from Equations 14 and 13) allow us to derive the following relationships in terms of relative demand and supply, so as to estimate the differential responses and elasticity of substitution across skill-levels. For $\Omega \equiv \sigma_E + \eta - \sigma_E \eta \gamma \psi_1$, we can derive:

$$log\frac{L_{sd}}{L_{ud}} = \frac{\sigma_E \eta}{\Omega}\left(log\frac{\theta_{sd}}{\theta_{ud}} - log\frac{\overline{V_{sd}}(H_{sd})}{\overline{V_{ud}}(H_{ud})} + \gamma\psi_2 logY_d + \gamma log\bar{Z}_d - \nu log\, hp_d\right) \tag{15}$$

$$log\frac{w_{sd}}{w_{ud}} = \frac{1}{\Omega}\left((\Omega-\eta)\, log\frac{\theta_{sd}}{\theta_{ud}} + \eta log\frac{\overline{V_{sd}}(H_{sd})}{\overline{V_{ud}}(H_{ud})} - \gamma\eta\psi_2\, logY_d - \gamma\eta log\bar{Z}_d + \nu\eta log\, hp_d\right) \tag{16}$$

Equation 15 makes an important point. The relationship between air quality $\bar{Z}_d$ and the relative supply of skilled and unskilled workers is not simply what one would think of as a partial migration response to pollution (captured by γ). Instead, in general equilibrium it is the result of future migration changes as wages change (given the size of η) in response to migration flows (given σ_E), changes to where production and pollution takes place as a response to where workers move to (given ψ_1 and ψ_2), and other factors (resulting changes to house prices, agglomeration, etc.). As such, the empirical results in Section 5 which estimate a relationship between migration and pollution, identify a coefficient that is a joint function of many parameters.

These are two important model results that we have already confirmed empirically. First, we have established that skilled labor falls as the amount of air pollution increases, consistent with Equation 15. Second, we have documented the result from Equation 16, which predicts that the returns to skill rise as pollution increases.

7.1 Estimating σ_E and γ

The parameters σ_E and γ, can be estimated with a few equations. The first is the equilibrium wage relationship:

$$log\frac{w_{sd}}{w_{ud}} = \frac{1}{\Omega}\left((\Omega-\eta)\, log\frac{\theta_{sd}}{\theta_{ud}} + \eta\lambda(Hukou_{sd} - Hukou_{ud}) - \gamma\eta\psi_2\, logY_d - \gamma\eta\, log\bar{Z}_d + \nu\eta log\, hp_d\right) + \varepsilon_{1d} \tag{16'}$$

We use pollution as a measure of $log\bar{Z}_d$. To estimate σ_E, we use the relative labor demand curve (Equation 13). As the relationship between the number of workers and wages is determined in equilibrium, we use a two-staged least squares (2SLS) relationship, where pollution is a shifter of labor supply. We rely on our empirical results that have established how pollution shifts the relative labor supply of skilled to unskilled workers. Since pollution shifts the labor supply curve is traces out the labor demand curve, letting us estimate an unbiased measure of the slope of the labor demand curve, which is σ_E. In other words, we estimate Equation 13 using 2SLS, where the first-stage is:

$$log\frac{L_{sd}}{L_{ud}} = \alpha_0 + \alpha_1\ log\ PM2.5_d + \varepsilon_{3d} \tag{17}$$

Table 3: Estimating Labor Demand Elasticities

	IV: Wind Direction and Coal Plants			
	Log(Wage Ratio)		Log (Skill Ratio)	
Log(PM 2.5)	0.971**	0.906**	-1.198**	-1.346**
	(0.377)	(0.413)	(0.589)	(0.624)
Cities	116	116	116	116
City Controls	Y	Y	Y	Y
Weight	Population 2015	Population 2000	Population 2015	Population 2000

City level regressions in 2015. Skilled workers denote those whose highest degree is some college or above, unskilled workers denote those whose highest degree is high school or bellow. City Controls include log of distance to Shanghai seaport, Tianjin seaport, and Shenzhen seaport. Standard errors clustered at the city level are reported in parentheses.

We perform such an exercise in Table 3. We estimate Equation 16 in the last 2 columns, where our outcome of interest is the stock of workers capturing net migration for all types of workers (whether or not they changed *hukou* location). Our IV estimates suggest that $\frac{\gamma\eta}{\Omega} = 0.9$. Together with the help of the first 2 columns, we can also estimate the relationship in Equation 13. To do this, we can simply take the ratio of the IV relationship for quantities of workers, and the income of workers. We find that $\frac{1}{\sigma_E} = 0.9/1.35 = 0.67$. This suggests that the elasticity of substitution across skill levels is 1.49, an estimate that is also close to the estimates found in the US (Card and Lemieux, 2001).

Notice, however, that while σ_E is identified by the change in relative wages in response to changes in relative skill-shares, γ is identified by the initial shift in relative skill-shares in response to pollution. From Equation 15, we can see that the coefficient on pollution shocks when the outcome is the relative skill-share, is a function of γ and other parameters, such that $\frac{\gamma\eta}{\Omega} = 1.35$ (from the last column of Table 3). This is because, even though γ determines the extent to which workers relocate when faced with higher pollution, this worker relocation affects a lot more in general equilibrium. As workers move, it affects wages (σ_E), which in turn affects future migration (η), and thereby affects where production and pollution is location (ψ_1 and ψ_2), which in turn affects migration responses, and so on. The advantage of the general equilibrium set up is that we can capture all these endogenous changes by estimating the necessary elasticity. Once we estimate the other parameters in the sections below, the supply response to pollution allows us to estimate $\gamma = -1.15$.

7.2 Estimating η:

Next, we estimate η. Most papers get the Frechet elasticity from the literature (for instance, Hsieh and Moretti (2018)). We estimate η is using an instrument that shifts out the relative demand curve (as that traces out the supply curve). Labor demand shocks change θ_{sd}, and shift out the relative labor demand curve. We leverage variation from city-specific differential trade shocks as our exogenous driver of labor demand shocks. Facchini et al. (2018) study how trade shocks in different parts of China affect internal migration. We follow a similar strategy to estimate the relationship in Equation 14.

We use variation coming from China's accession to the WTO in 2001, and the subsequent change in city-level growth to identify the effects of demand shocks on population mobility. This is the same strategy used by researchers to study the impact of Chinese exports on US manufacturing employment (Pierce and Schott, 2016). Prior to joining the WTO, the US gave China lower Normal Trade Relations (NTR) tariffs, which were continually renewed by Congress. Joining the WTO reduces the renewal uncertainty, captured by the NTR Gap, defined to be the difference between the non NTR tariff and the NTR tariff.

We create city-level uncertainty measured by looking at the weighted sum of industry employment shares, interacted with the NTR gaps:

$$Trade\ Shock_d = \sum_i \frac{L_{di}^{1998-2000}}{\sum_j L_{dj}^{1998-2000}} \times (NTR\ tariff_i - nonNTR\ tariff_i) \tag{18}$$

Intuitively, the skill-biased trade shock will raise the demand for some occupations more

Table 4: Estimating Labor Supply Elasticities

	Log(Export Growth 2001-13)		Log(Skill-Wage Ratio)		Log(Skill-Emp Ratio)	
NTR Gap Instrument	5.592***	5.691***				
	(1.731)	(1.711)				
Log(Export Growth)			0.295*	0.307**	0.0474**	0.0498**
			(0.151)	(0.153)	(0.0212)	(0.0219)
Observations	287	287	149	149	283	283
Controls	No	Yes	No	Yes	No	Yes
First Stage F	10.43	11.07				

The NTR gap is measured as the gap in Normal Trade Relation (NTR) tariffs and the non-NTR tariffs. We weight the product level tariffs by city-level production shares. Skill-wage ratio denotes the wage ratio of workers whose highest degree is some college or above to those whose highest degree is high school or bellow. Skill-employment ratio denotes the ratio of workers whose highest degree is some college or above to those whose highest degree is high school or below. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

than others. This changes the relative wages by city in response to the trade shock. If skilled wages rise more than unskilled wages in a city, then that city will attract relatively more skilled migrant workers. This helps us identify the relative labor supply curve in response to changes in wages. Table 4 allows us to estimate η from Equation 14 leveraging variation that arises from China's accession to the WTO. Our estimates suggest that $\eta = 0.049/0.31 = 0.16$. Note that as workers move in response to higher wages, this will affect where production takes place, and as a result, where pollution is located (based on ψ_1 and ψ_2). So our model and estimation allows for the fact that trade shocks will also affect the amount of pollution via production and migration responses.

7.3 Estimating λ:

Hukou restrictions affect the ease with which skilled and unskilled workers can relocate, and access jobs and public services in different destination cities. Such restrictions make it relatively more difficult for unskilled workers to move, and as such, could be a source of labor misallocation. If, for instance, the skilled workers leave a polluted city, but unskilled workers find it difficult to leave with them, then given the complementarity between the skilled and unskilled, there will be a fall in economic output. Bringing the skilled workers back would raise output as the skilled would be more productive working with the unskilled, and the unskilled would be more productive working with the skilled. Easing migration restrictions would mean that unskilled workers can relocate with skilled workers and take advantage of the complementarity in production between the skilled and unskilled.

Table 5: Hukou Restrictions and Skill Ratio

	Log (Skill Ratio)			
	Full sample		Hukou > 0	
Hukou Index	2.474***	2.335***	1.971***	1.811***
	(0.176)	(0.257)	(0.190)	(0.263)
Observations	232	232	102	102
Adjusted R-squared	0.312	0.330	0.224	0.253
City Controls	N	Y	N	Y

The hukou index is a measurement of the difficulty to get local urban hukou across different criteria, including investments and job status. The higher the index is, the more restrictive the hukou system is. City level regressions in 2015. Skilled workers denote those whose highest degree is some college or above, unskilled workers denote those whose highest degree is high school or bellow. City Controls include log of distance to Shanghai seaport, Tianjin seaport, and Shenzhen seaport. Standard errors clustered at the city level are reported in parentheses.

In Table 5 we estimate the relationship between the restrictiveness of *hukou* policy and the skill ratio. We use data on the *hukou* index, which is a measure created by Zhang et al. (2018). The *hukou* index determines the ease with which skilled and unskilled workers can move to different cities, which in turn, may affect city-level amenities and air quality (as captured by the ψ parameters). To be specific, the index measures the difficulty of getting a local *hukou* based on job status, family reunion motives, local investments, and contribution to the city workforce. A higher index determines a more restrictive policy. The index is highest for Beijing, followed by many of the other major cities.

The results in Table 5 indicate that, as expected, the more restrictive the *hukou* policy is, the relatively more difficult it is for low skill workers to move to those cities. As such, a higher *hukou* index leads to a larger skill-ratio in destination cities. Traversing the inter-quartile range of the *hukou* index (a change of 0.029 units) raises the skill-ratio by 5%. This is a modest, but meaningful change in the skill-ratio. The modest magnitude suggests that there are other factors that also drive changes in the skill ratio. Based on Equation 15 and the last column in Table 5 we know that $\lambda = -1.811 \times (\Omega/\sigma_E \times \eta) = -17$.

7.4 Estimating ψ parameters:

The remaining parameters include how changes in the skill-ratio affect the amount of pollution in each city and house prices and determine agglomeration/congestion effects. In order to estimate ψ_1 we require variation in the skill-ratio that is not driven by the air quality. We derive this variation from the *hukou* policy.

In Table 6 we see that the *hukou* index is a strong predictor of the skill ratio, and the skill ratio does indeed affect PM 2.5. We find that $\psi_1 = -0.137$.

Table 6: How Skill-distribution Affects Air Quality

	Ln Skill Ratio				LnPM2.5	
	Full Sample		Hukou Index> 0		Full Sample	Hukou Index> 0
Hukou Index	1.880***	1.021***	1.507***	0.792***		
	(0.221)	(0.190)	(0.281)	(0.213)		
Ln Skill Ratio					-0.154***	-0.137***
					(0.0380)	(0.0420)
Observations	376	376	212	212	376	212
Adjusted R-squared	0.145	0.474	0.091	0.523	0.366	0.346
City Controls		Y		Y	Y	Y

We use the Hukou Index as an instrument for skill-ratio, and examine how changes in the skill ratio affect air quality. We do our analysis for two samples: the full sample of cities, and those with a hukou index. In the full sample, we assign a value of 0 for those cities with no hukou index. City controls include log of the distance to the three nearest seaports. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

We still have to estimate how changes in local GDP affect air quality, captured by parameter ψ_2. $\bar{A}_d$ captures features of the local area (land, etc.) and indsutry-specific TFP growth. We use an industry-level shift-share instrument to leverage variation in GDP. Differential growth in technological changes across industries will cause some Chinese cities to grow faster than others, given the baseline distribution of industries across regions. We study how this source of variation drives changes in GDP, which in turn affects PM 2.5. We use the 2-digit industrial classification for manufacturing industries, and aggregate the firm-level data to the city-industry level at first to create a measure of the value of production in industry i and city c: $production\ value_{c,i}$. We then estimate the industry-level growth rate in all the other cities, $growth_{-c,i}$ between 1998 and 2013. We finally create the shift-share instrument at the city level:

$$Shift\ share_c = \sum_i \left(\frac{production\ value_{c,i}}{\sum_i production\ value_{c,i}} \times growth_{-c,i} \right) \tag{19}$$

Table 7 describes this relationship. We find that the shift-share is a strong predictor of city-level GDP, as evidenced by the first stage. Our two-staged least squares estimated tell us that $\psi_2 = 0.358$.

Table 7: How Production Affects Air Quality

	First Stage Ln GDP		OLS Ln PM2.5		Bartik IV Ln PM2.5	
Bartik IV	0.366***	0.249***				
	(0.0409)	(0.0250)				
Log(GDP)			0.194***	0.0990*	0.498***	0.358***
			(0.0373)	(0.0521)	(0.107)	(0.138)
Observations	245	245	257	257	245	245
Adjusted R-squared	0.368	0.725	0.127	0.741	0.00	0.662
City Controls	N	Y	N	Y	N	Y
Province Fixed Effect	N	Y	N	Y	N	Y

We use the Bartik for industry-level growth as an instrument for GDP, and examine how changes in GDP affect air quality. City controls include log of the distance to the three nearest seaports. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

Finally, in an extension of our exercise, we consider how house price changes may affect our estimates. To do so, we need to estimate ψ_4, the elasticity of house prices with respect to the share of skilled workers. To do this, we compile rental price data from the China mini census of 2005. To identify the effects of changes in skill ratio on house prices, we leverage the higher education expansion in China in the early 2000s.[21] Table B5 establishes the relationship between changes in the skill ratio and house prices, an elasticity of $\psi_4 = 1.238$.

[21] In January 1999 the Ministry of Education (MOE) announced an admission plan of 1.3 million for three and four-year college programs, a 20% increase over 1998. The following June it revised the admission plan to 1.56 million, an unprecedented increase of 44% over the previous year. College admissions grew annually by more than 40% in both 1999 and 2000, and by about 20% over the next five years. The gross college enrollment rate among 18–22 year-olds increased from 9.8% in 1998 to 24.2% in 2009. The year 2003 saw the first flow of four-year graduates into the job market as a result of the expansion. The number of students graduating from regular higher education institutions was 2.12 million, a 46.2% increase from the previous year. Che and Zhang (2018) use this policy and establish this to be a credible identification strategy by showing that the higher education expansion policy was not targeted at any sector.

8 Growth Accounting: GDP, and TFP

Output in city d depends on the set of parameters: $\{\theta_{sd}, \sigma_E, \eta, \phi_1, \phi_2, \psi_1, \psi_2, \psi_4, \lambda\}$, and a set of quantities: $\{A_d, L_{sd}, L_{ud}, H_{sd}\}$. Notice, however, that L_{sd} is an 'endogenous' quantity as it is determined by migration decisions. Based on Equation 15 above, it is clear that: $L_{sd} = f(L_{ud}; \theta_{sd}, \gamma_s, \sigma_E, \eta, \phi_1, \phi_2, \psi_1, \psi_2, \psi_4, \lambda)$. Similarly, we know that $A_d = \bar{A}_d Z_d^{\phi_1} \frac{L_{sd}}{L_{ud}}^{\phi_2}$. This implies that we can re-write, output in city d, to be a function of the parameters and 'exogenous' quantities: $\{L_{ud}, Z_d, \bar{A}_d, H_{sd}\}$. Total output in the country is simply the sum of all regional outputs:

$$Y = \sum_{d}^{D} Y_d \left(\bar{A}_d, L_{ud}, Z_d, H_{sd}\ ;\ \theta_{sd}, \sigma_E, \gamma_s, \eta, \phi_1, \phi_2, \psi_1, \psi_2, \psi_4, \lambda\right) \tag{20}$$

Changes in pollution levels Z_d will affect the location of workers and TFP, thereby changing output Y_d in the city, and in other cities as well.

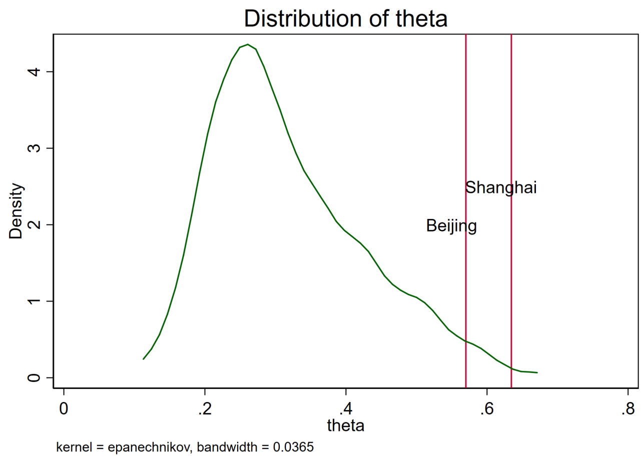
8.1 Measuring θ_{sd}

We measure θ_{sd} from data on labor shares in the wage bill and the properties of a CES function. θ_{sd} varies at the city level by the amount of skill-biased capital in each city. We use the following relationship, and information on wages and number of workers to measure θ_{sd} at the city level:

$$\frac{w_{sd}L_{sd}}{w_{sd}L_{sd} + w_{ud}L_{ud}} = \frac{\theta_{sd}L_{sd}^{\frac{\sigma_E-1}{\sigma_E}}}{\theta_{sd}L_{sd}^{\frac{\sigma_E-1}{\sigma_E}} + (1-\theta_{sd})L_{ud}^{\frac{\sigma_E-1}{\sigma_E}}} \tag{21}$$

We plot the city-level distribution of θ_{sd} in Figure 10. Beijing and Shanghai have the highest amount of skill-biased capital in the country, which almost double the amount of the median city, many of which lie in rural areas.

Figure 10: Distribution of θ_{sd} across cities, from Equation 21



8.2 TFP, Agglomeration and the Effect of Pollution on Productivity

To quantify any changes to output, we first need a measure of TFP $\bar{A}_d$. Like the rest of the literature, TFP is measured as the city-level aggregate residual (not the relative skilled-unskilled residual). After accounting for the optimal (unbiased) capital flows, output is simply: $Y_d = A_d \left(\frac{1-\varrho}{R^*}\right)^{\frac{1-\varrho}{\varrho}} L_d$. We use our wage data to determine that the share of labor in GDP is on average $\varrho = 0.29$. We then estimate Equation 22 in the first column of Table 8, where $\log \bar{\varrho} \equiv \left[\left(\frac{1-\varrho}{\varrho}\right) \log\left(\frac{1-\varrho}{R^*}\right)\right]$ is common across all cities and workers. ϱ is nothing but the labor share of income in the city. We verify that the coefficient on $Log(L_d)$ is not statistically distinguishable from 1.

$$\log Y_d = \log \bar{\varrho} + \log L_d + \log A_d \ , \tag{22}$$

All other parts of $L_d = f(L_{ud}; \theta_{sd}, \sigma_E, \eta, \psi_1, \psi_2, \psi_4, \lambda, \nu)$ are now observable. So we measure $\log A_d$ as a residual from our estimation of Equation 22 in the first column of Table 8. We then use Equation 23 to study how our measure of TFP correlates with pollution as instrumented with the coal-fired plants and wind direction instrument.

$$\log A_d = \log \bar{A}_d + \phi_1 \log Z_d + \phi_2 \log \frac{L_{sd}}{L_{ud}} \ , \tag{23}$$

where ϕ_1 is the elasticity of pollution with aggregate TFP.[22] Notice, A_d may capture other

[22]TFP may also be positively associated with pollution. This may generally be true if more production leads

Table 8: Pollution, Skill Ratios and TFP

	Log(GDP)	OLS Log(TFP)	IV Log(TFP)	IV Log(TFP)
$Log(L_d)$	1.077***			
	(0.0566)			
$Log(PM2.5)$		-0.000406	-0.0963	-0.0399
		(0.108)	(0.169)	(0.109)
$Log(L_s/L_u)$				0.116***
				(0.0389)
Observations	232	232	232	110
F-stat			22.8	26.6
Year Fixed Effect	Yes	Yes	Yes	No
Province Trends	Yes	Yes	Yes	No

Notes: The residuals from the regression in the first column are used to calculate TFP. The latter columns estimate the relationship between PM 2.5 and TFP. We use the coal-fired power plants instrument in the IV specification for column 3. In column 4 we also look at the effect of the skill ratio on TFP, where we instrument for the skill ratio using the higher education expansion instrument.

drivers of city-level TFP, like land, housing supply, innovation. Equation 23 also allows the skill-ratio to directly affect the amount of TFP in a city. If there are Arrow (1962) style innovation spillovers, it would be captured by the agglomeration elasticity ϕ_2.[23]

In Table 8 we estimate ϕ_1 using Equation 23, and leveraging our instrumental variables strategy for PM 2.5 emissions. While not precisely estimated, we conclude that $\phi_1 = -0.096$. (Chang et al., 2019) document an elasticity of -0.023 in the context of call-center workers in China, while (Adhvaryu et al., 2016) estimate an elasticity of -0.052 for garment sector workers in India.

In addition, we estimate estimate the impact of a changing skill-ratio on TFP to capture in agglomeration or congestion effects. Here we instrument for the skill ratio using the higher education expansion in China over the early 2000s. We aggregate the number of college graduates by city for the first graduating cohort of the expansion (2001-2004). Our estimate of ϕ_2 suggest a meaningful agglomeration effect. Work by Zhang and Yao (2010) estimate an agglomeration elasticity of 0.02, which is it at the lower end of our estimated range.

In Table 9 we summarize all these parameter values and sources of variation. Given our

to more pollution, but we are using an IV for pollution here.

[23]Agglomeration in this model is represented by the production of non-excludable ideas. As such, innovators are not directly compensated for the ideas, but instead overall output increases, benefiting all in the local economy.

Table 9: Summary of Parameter Estimation

Parameter	Value	Definition	Identifying Variation
σ_E	1.49	Skill-elasticity of substitution Relative labor demand elasticity	Pollution-driven geographic sorting Changes in skill ratio affect wages
η	0.16	Relative labor supply elasticity	Trade shocks post WTO
γ	-1.15	Compensating differential elasticity	Pollution-driven response To skill ratios
θ_{sd}	[0.15,0.64]	Skill-biased capital	Skill-labor share in wage bill
ψ_1	-0.137	Pollution response to changing skill	Hukou index policy
ψ_2	0.358	Pollution response to GDP	Industry-level TFP growth
ψ_4	1.238	House price response to skill ratio	University expansion
ϕ_1	-0.0963	TFP response to pollution	Pollution IV affect TFP residual
ϕ_2	0.02,0.116	TFP response to skill ratio	University expansion
λ	-17	Skill ratio response to hukou index	Hukou index

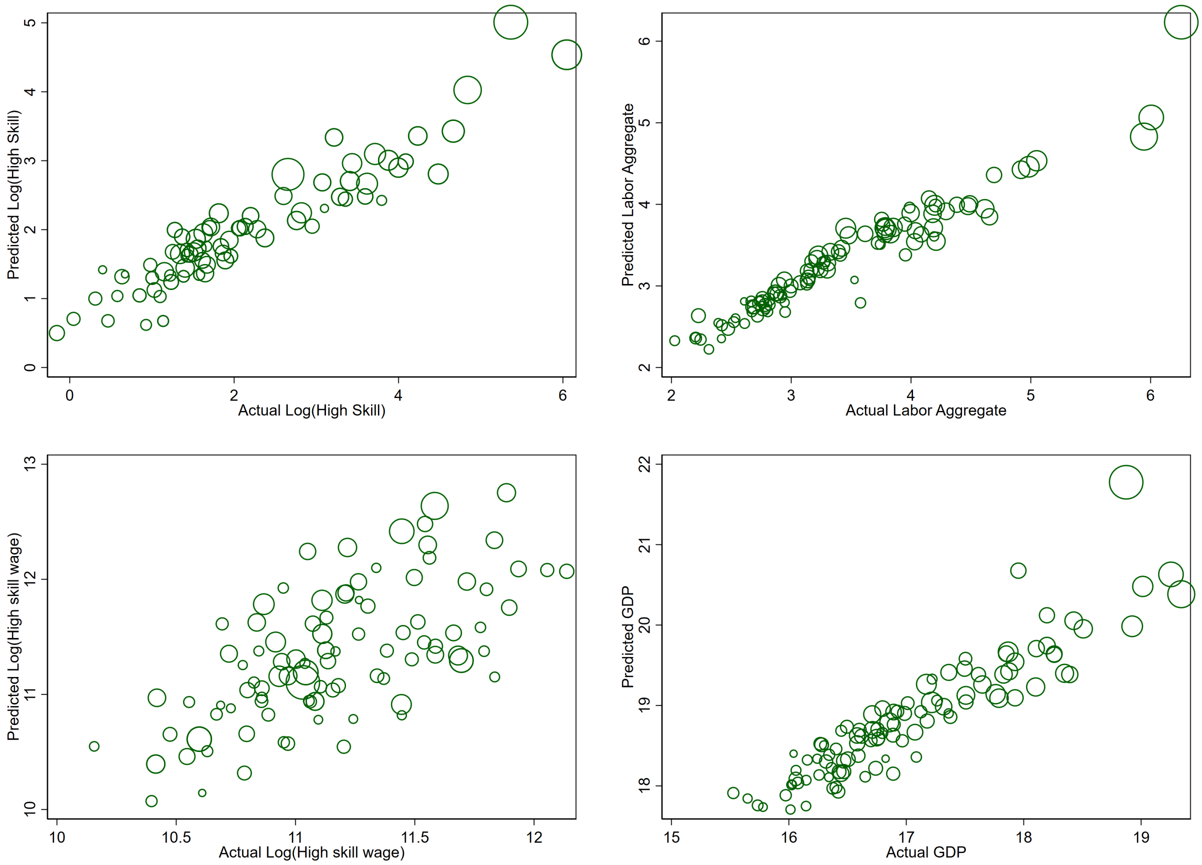
Notes: We summarize the parameter estimation using different instruments in this table. The values are from Tables 3-8, Table B5 and Figure 10. The top half of our table lists the primary parameters for our main model. The lower half of the table includes the additional parameters that complete the estimation process.

estimated parameters that determine GDP, Y_d, we can create a model-predicted measure of GDP based on our estimates.

8.3 Validation Exercises

There are a few tests that help validate the assumptions underlying this model. We test for model fit at the city level in Figure 11. First, we test model-fit by plotting the predicted number of high skill workers against the actual high-skill workforce in a city, and the predicted labor aggregate against the actual labor aggregate. In the lower panels we first plot actual and predicted skill-wages. Notice, these are not necessarily an out-of-sample test as we these data when estimating the different parameters of the model. Last, we plot city-level GDP Y_d against actual city-level GDP. To calculate the predicted values across cities, we rely on the estimated parameters of the model. For instance, the prediction of the log of the number of high-skill workers, labor aggregate and wages depend on $(L_{ud}; \theta_{sd}, \sigma_E, \eta, \psi_1, \psi_2, \psi_4, \lambda, \nu)$, whereas the

Figure 11: Model Fit and Validation in 2015



Notes: We plot the actual and predicted relationship between our main variables, where the predictions are based on model-estimated parameters.

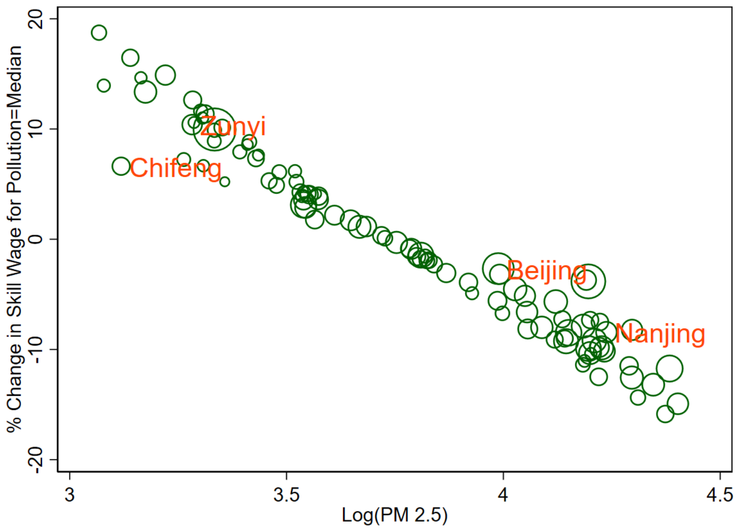
prediction for GDP additionally depends on $(\bar{A}_d, \phi_1, \phi_2)$.

9 Counterfactuals: The Gains from Re-Allocation

To quantify the role played by poor air quality in determining the productivity losses via migration, we do a few counterfactual exercises. First, we quantify how much of the wage gap across regions is because of pollution. We conduct an exercise where we change the amount of pollution in a city to be that of the median city in the country. This means raising pollution levels in low polluting cities, and lowering them in high polluting cities. In Figure we show how wages would change with this reallocation of pollution. As pollution is lowered in cities like Beijing and Nanjing, there is an inflow of skilled workers that lowers the skilled wage. The change in wage is the same as the (endogenously determined) compensating differential. Such an exercise also allows us to quantify how much of the wage gap across cities is due to pollution.

In the context of this exercise, the skilled wage gap between Beijing and Zunyi is bridged by 5.5%, and the gap between Nanjing and Chifeng by about 8.8%.

Figure 12: Explaining the Wage Gap



Notes: We plot the change in the skilled wage when the amount of pollution in the city is changed to be equal to the pollution in the median city. The horizontal axis plots the baseline amount of pollution in a city. The vertical axis plots the change in the skilled wage as this baseline pollution is changed to be that of the median city.

Second, in Table 10 we examine what would happen if the city with the highest levels of pollution (Tianjin) reduced PM 2.5 to be the same as in the median Chinese city. GDP would rise by 4.82% simply due to high-skill workers now moving to Tianjin. If we allow for the fact that reducing pollution also has health benefits, and more high-skill workers means more positive agglomeration spillovers, GDP would rise by 10.17%. If, in addition to reducing PM 2.5, we also relax hukou restrictions in Tianjin to be the same as the 25th percentile of cities with hukou (i.e. of the big hukou cities, Tianjin is assigned the value of the 25th percentile city), then GDP would rise by 7.11% simply due to reallocation of workers, and by 12.62% given the other changes to TFP. As a comparison, if we only relaxed hukou without reducing pollution, the rise in GDP would be smaller.

Finally, in Table 11 we relocate where pollution takes place. We move pollution from cities that have more skill-biased capital (as measured by baseline skill-shares θ_{sd}) to those with less skill-biased capital, while still keeping the mean amount of pollution in the country the same as before.[24] Doing so raises GDP by 3.86% simply due to reallocation, and by 5.99% after taking

[24]Specifically, we remove pollution from places with high θ_{sd} and allocate it to places with low θ_{sd} by doing:

Table 10: The Effect of Reducing Pollution in One City

	Change in Tianjin GDP (%)	
Change to Tianjin only:	Due to labor re-allocation	Labor reallocation and TFP
Reduce PM2.5 to median city	4.82	10.17
Reduce PM2.5 and relax hukou	7.11	12.62
Relax hukou only	2.52	5.05

Notes: In this counterfactual exercise we reduce the amount of pollution in Tianjin to be the pollution level of the median city in China (row 1). We then reduce pollution and relax the hukou restriction (row 2) to be the same hukou restriction as the 25th percentile of the large cities with hukou restrictions. Finally (row 3) we relax the hukou to the same as the 25th percentile of the large cities in China without reducing pollution. Column 1 shows the gain to GDP due to the re-allocation of labor channel only. Column 2 shows the gain to Tianjin GDP after also accounting for changes in TFP due to the direct effect of pollution on health/productivity, and changes in agglomeration/congestion.

into account the other changes to TFP. Note here that the agglomeration and health-benefits forces may act in opposite directions. If we also relax *hukou* restrictions in high-skill share cities, while making low skill share cities more restrictive (keeping the mean hukou restrictions the same), GDP rises by 18.8% due to reallocation of workers, and by 27.8% including the other TFP effects. As a benchmark, if we were only to re-allocate hukou restrictions the rise in GDP would be 6.25% and 6.89% respectively.

10 Discussion

Our analysis highlights a simple fact: geographic sorting based on differences in amenity values, that leads to differences in skill-distributions across regions will cause a spatial sorting on skill. In the context of China, we find that skilled workers leave places with higher pollution and move to places with better air quality. Our quantitative exercise estimates how much productivity would rise by if we lowered pollution in major cities, and as such moved skilled workers from areas of low wages to areas that have higher returns to skill. We find that in China, poor air quality substantially contributes to moving college educated workers to areas with better air quality but lower marginal products. As such, reducing pollution levels efficiently reallocate

$Log(PM2.5) - (\theta_{sd} - \overline{\theta_{sd}})$.

Table 11: The Effect of Changing Pollution in All Cities by Skill Biased Capital

	Change in China GDP (%)	
Change:	Due to labor re-allocation	Labor reallocation and TFP
Change PM 2.5 by skill share	3.86	5.99
Change PM2.5 and relax hukou	18.8	27.8
Relax hukou only	6.25	6.89

Notes: In this counterfactual exercise we change the amount of pollution in each city based on the baseline skill-share (row 1), which captures the amount of skill biased capital. As such, low skill-biased capital receive more pollution and high skill-share cities see a reduction in pollution so as to keep the mean amount of pollution the same. We then change pollution and change the hukou restriction (row 2) in all cities based on baseline skill-share. Now, low skill share cities have more restrictive hukou and high skill share cities have less restrictive hukou. Finally (row 3) we change the hukou only without changing pollution. Column 1 shows the gain to GDP due to the re-allocation of labor channel only. Column 2 shows the gain to GDP after also accounting for changes in TFP due to the direct effect of pollution on health/productivity, and changes in agglomeration/congestion.

workers and raises aggregate output.

These results neatly bridge both the literature on how increasing mobility in developing countries can substantially produce large income gains (Bryan et al., 2014), and how a more efficient allocation of factor inputs will increase aggregate productivity (Hsieh and Klenow, 2009). As industrialization in many developing countries worsens air quality, our results suggest that the movement of people across cities in response to such disamenities will have indirect consequences on longer term growth and economic development.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2016). Management and shocks to worker productivity. *International Growth Center (IGC)*. Working Paper.

Andreoni, J. and A. Levinson (2001). The simple analytics of the environmental kuznetz curve. *Journal of Public Economics 80*(2), 269–286.

Arceo, E., R. Hanna, and P. Oliva (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal 126*(591), 257–280.

Arrow, K. J. (1962). Economic Welfare and the Allocation of Resources for Innovation. *NBER working paper c2144*.

Au, C. and V. Henderson (2006). How Migration Restrictions Limit Agglomeration and Productivity in China. *Journal of Development Economics 80*(2), 350–388.

Bayer, P., N. Keohane, and C. Timmins (2009). Migration and Hedonic Value: The Case of Air Quality. *Journal of Environmental Resources and Management 58*(1), 1–14.

Bazzi, S. (2017). Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics 9*(2), 219–55.

Brandt, L., C.-T. Hsieh, and X. Zhu (2008). Growth and structural transformation in china. *China's great economic transformation*, 683–728.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica 82*(5), 1671–1748.

Bryan, G. and M. Morten (2018). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*.

Card, D. and T. Lemieux (2001, May). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics 116*(2), 705–746.

Chang, T., J. G. Zivin, T. Gross, and M. Neidell (2019). The effect of pollution on worker productivity: Evidence from call center workers in china. *American Economic Journal: Applied Economics 11*(1), 151–172.

Che, Y. and L. Zhang (2018). Human capital, technology adoption and firm performance: Impacts of china's higher education expansion in the late 1990s. *Economic Journal 128*(614), 2282–2320.

Chen, S., P. Oliva, and P. Zhang (2017). The effect of air pollution on migration: Evidence from china. *NBER Working Papers*.

Chen, Y., A. Ebenstein, M. Greenstone, and H. 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 of the United States of America 110*(32), 12936–12941.

Chen, Y., G. Z. Jin, N. Kumar, and G. Shi (2012). Gaming in air pollution data? lessons from china. *The BE Journal of Economic Analysis & Policy 12*(3).

Clemens, M. A., C. E. Montenegro, and L. Pritchett (2019). The place premium: Bounding the price equivalent of migration barriers. *Review of Economics and Statistics 101*(2), 201–213.

Combes, P.-P., S. Demurger, S. Li, and J. Wang (2019). Unequal migration and urbanisation gains in China. *Journal of Development Economics*.

Eaton, J. and S. Kortum (2002). Technology, Geography and Trade. *Econometrica 70*, 1741–1779.

Ebenstein, A., M. Fan, M. Greenstone, G. He, P. Yin, and M. Zhou (2015). Growth, pollution, and life expectancy: China from 1991-2012. *American Economic Review 105*(5), 226–31.

Ebenstein, A., M. Fan, M. Greenstone, G. He, and M. Zhou (2017). New 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 114*(39).

Facchini, G., M. Y. Liu, A. M. Mayda, and M. Zhou (2018). China's "great migration": The impact of the reduction in trade policy uncertainty.

Freeman, R., W. Liang, R. Song, and C. Timmins (2019). Willingness to pay for clean air in china. *Journal of Environmental Economics and Management*.

Ghanem, D. and J. Zhang (2014). Effortless perfection: Do chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management 68*(2), 203–225.

Gollin, D., D. Lagakos, and M. Waugh (2014). The agricultural productivity gap. *Quarterly Journal of Economics*, 939–993.

Heblich, S., A. Trew, and Y. Zylberberg (2019). East side story: Historic pollution and neighborhood segregation. *SERC Discussion Paper 208*.

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

Hsieh, C. T. and P. Klenow (2009). Misallocation and manufacturing TFP in china and india. *The Quarterly Journal of Economics* (4). November.

Hsieh, C. T. and E. Moretti (2018). Housing constraints and spatial misallocation. *Mimeo*.

Ito, K. and S. Zhang (2019). Willingness to pay for clean air: Evidence from air purifier markets in china. *Journal of Political Economy*.

Jans, J., P. Johansson, and P. Nilsson (2014). Economic status, air quality, and child health: Evidence from inversion episodes. *Working Paper*.

Kapur, M. (2019, Nov). Many Delhi Residents Want to Escape the Smog. But Where Can They Go? *Quartz India*.

Kone, Z., M. Liu, A. Mattoo, C. Ozden, and S. Sharma (2018). Internal Borders and Migration in India. *Journal of Economic Geography 18*(4), 729–759.

Lagakos, D., M. Mobarak, and M. Waugh (2019). The welfare effects of encouraging rural-urban migration. *Working Paper*.

Lewis, W. A. (1954). Economic development with unlimited supplies of labor. *The Manchester School*. May.

Moretti, E. (2011). Local labor markets. *Handbook of Labor Economics 4b*.

Munshi, K. and M. Rosenzweig (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review 106*(1), 46–98.

Pierce, J. R. and P. K. Schott (2016). The surprisingly swift decline of u.s. manufacturing employment. *American Economic Review 106*(7), 1632–62.

Restuccia, D. and R. Rogerson (2013). Misallocation and productivity. *Review of Economic dynamics 16*(1), 1–10.

Restuccia, D. and R. Rogerson (2017). The causes and costs of misallocation. *Journal of Economic Perspectives 31*(3), 151–74.

Sharma, K. and H. Chandna (2019, Nov). How Air Pollution Has Become a Big Factor in Indians' Decision to Work in delhi. *The Print*.

Tombe, T. and X. Zhu (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review 109*(5), 1843–72.

Van Donkelaar, A., R. Martin, M. Brauer, N. C. Hsu, R. Kahn, R. Levy, A. Lyapustin, A. Sayer, and D. Winker (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental Science & Technology 50*(7), 3762–3772.

Young, A. (2014). Structural transformation, the mismeasurement of productivity growth, and the cost disease of services. *American Economic Review 104*(11), 3635–67.

Zhang, H. and X. Yao (2010). Economic agglomeration, externalities and corporate labor productivity: Evidence from the zhejiang province. *Management World* (12).

Zhang, J., R. Wang, and C. Lu (2018). A Quantitative Analysis of Hukou Reform in Chinese Cities: 2000-2016. *Growth and Change 50*(1), 201–221.

Zheng, S. and M. E. Kahn (2013). Understanding china's urban pollution dynamics. *Journal of Economic Literature 51*(3), 731–72.

Zivin, J. G. and M. Neidell (2012). The impact of pollution on worker productivity. *American Economic Review 102*(7).

A Appendix: Robustness of Empirical Specifications

We conduct a few meaningful robustness checks to evaluate the concreteness of the empirical relationship between air quality and migration decisions. We explore threats to identification, alternative model specifications, and different instrumental variables.

A.1 Alternative Instruments and Sources of Variation

First, we explore a few different sources of underlying variation. We check our results in simple OLS specifications, before trying two different instrumental variables strategies discussed in recent work on China. Specifically, we study the variation underlying the instrument based on wind direction and coal-fired power plants (Freeman, Liang, Song, and Timmins, 2019), and the variation in air quality driven by the number and strength of thermal inversions (Chen, Oliva, and Zhang, 2017).

Then, we explore threats to identification for our two main instruments. Here, we test concerns of the endogenous placements of power plants, whereby policy makers may use the same function – the simultaneous interaction between wind directions, distance to cities and coal consumption – to determine where to place new plants. We show that the IV is not predicted by baseline city-level characteristics, and we exclude any plants built within different distance radii around the city. We may still think that *newly* built plants are endogenously placed. Yet, our results are robust to relying on old power plants, and to make it even more conservative the newly built plants are in the 'control' group. Last, we exclude major cities and provincial capitals that may have undue influence in this process.

We also explore the variation underlying the thermal inversions IV. We fail to find meaningful predictors of future inversions, and as such conclude that such events are random.

In the section that follows thereafter, we explore the time dimension of our data in various panel specifications. Since migration decisions are important and long-lasting, we do not expect short-term shocks to influence such long-term decisions. As such, most of the underlying variation we rely on is based on changes in the cross-section. Yet, we explore how persistent shocks over a long period of time can affect migration decisions.

Finally, we explore the variation generated by China's Huai river heating policy (Chen et al., 2013), but for once fail to find substantial effects on out-migration rates, even as they do help predict differential in-migration.

A.1.1 Instruments Variable Estimates

In our main text, we show how in Table 1 there is a strong relationship between out-migration and poor air quality. However, we expect an OLS relationship to also pick up systematic underlying differences in the industrial composition of cities. To purge such underlying differences from our estimates, we explore a few other empirical strategies. In Table A1 we show the strength of the first stage relationships between our different instruments and our independent variable of interest.

Table A1: The First Stage Across Different Instruments

	Log(PM 2.5)	Log(PM 2.5)	Log(PM 2.5)
Wind Direction and Coal Plants	8.57e-07***		
	(1.81e-07)		
Number of inversions		0.00116***	
		(0.000204)	
Strength of inversions			0.000311***
			(9.89e-05)
F-value in the 1st Stage	22.42	32.33	9.91
Observations	285	285	285
Adjusted R-squared	0.403	0.422	0.377
City Controls	Y	Y	Y
Demographics	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

In Table A2 we study how variation in PM 2.5 from our instrument of the interaction between coal-fired power plants, distance to plants and wind direction affects migration probabilities. We include different sets of controls, including distance to major industrial centers and big cities to account for the fact that being close to cities lowers migration costs. In our most conservative specification, for higher educated workers, we estimate a magnitude similar to that of our OLS specification – here, a 10% increase in PM 2.5 leads to a 0.696 percentage point increase in out-migration rates for those with a college degree. The effects on low skilled workers is much

Table A2: Wind Direction, Distance and Coal-fired Power Plants

	Dep var: Leave hukou city indicator					
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0363	0.0196	0.0696*	0.0543	0.0302	0.105**
	(0.0484)	(0.0568)	(0.0383)	(0.0502)	(0.0596)	(0.0433)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.031	0.029	0.084	0.030	0.029	0.076
Distance to ports	Y	Y	Y	Y	Y	Y
Distance to cities	Y	Y	Y	N	N	N
Demographics	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

smaller.

In Table A3 we leverage the variation coming from thermal inversions over the preceding ten years (2005-2015). In both sets of results we find meaningful relationships between PM 2.5 and out-migration rates. The strength of inversions regressions suggest that a 10% increase in PM 2.5 raises out-migration rates by 1.3 percentage points for those with a college degree.

Finally, in Table A4 we combine the different instruments and perform simple Hansen overidentification tests. We report both the J statistics and p-values of these tests in the tables. Using the combined instruments produce similar results on migration rates.

Table A3: IV: Thermal Inversions

	IV: Number of Thermal Inversions			IV: Strength of Thermal Inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0910**	0.0923**	0.142***	0.0485	0.0519	0.134***
	(0.0360)	(0.0398)	(0.0331)	(0.0428)	(0.0467)	(0.0495)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.030	0.028	0.079	0.032	0.030	0.080
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the number of thermal inversions in the last ten years. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A4: Combined Instruments

Instruments:	Wind and Coal + no. of inversions			Wind and Coal + strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0684*	0.0646	0.107***	0.0413	0.0341	0.0891***
	(0.0363)	(0.0406)	(0.0282)	(0.0402)	(0.0452)	(0.0325)
Adjusted R-squared	0.031	0.030	0.083	0.032	0.030	0.084
Hansen J statistic	2.019	2.828	2.83	0.072	0.403	1.427
Hansen pval	0.1553	0.0926	0.0925	0.7881	0.5254	0.2322
Observations	444,490	368,957	75,533	444,490	368,957	75,533
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

While our earlier results examine the out-migration decision of individuals, in Tables A5 and A6 we examine the destination choices across various instrumental variables strategies. We find that even for in-migration decisions, the responsive of high-skill workers is greater than that of low-skill workers.

Table A5: Wind Direction IV: In Migration and Destination PM 2.5

	Dep var: Share of in-migrants		
	Full Sample	Low edu	High edu
Log(PM 2.5)	-0.618	-0.592	-0.958**
	(0.442)	(0.478)	(0.430)
Observations	290	290	290
Adjusted R-squared	0.341	0.313	0.327
City Controls	Y	Y	Y

Notes: Dependent variable is share of group that are immigrants. Independent variable is destination city PM 2.5. Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A6: Thermal Inversions IV: In Migration and Destination PM 2.5

	Dep var: Share of in-migrants IV: Number of Thermal Inversions					
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM2.5: 1998-2015)	-0.601**	-0.511	-1.036***			
	(0.301)	(0.315)	(0.358)			
Ln(PM2.5: 2001-2015)				-0.587**	-0.498	-1.019***
				(0.295)	(0.309)	(0.350)
Constant	1.187	0.144	3.135	1.210	0.156	3.228
	(2.354)	(2.464)	(2.797)	(2.372)	(2.486)	(2.815)
Observations	290	290	290	290	290	290
Adjusted R-squared	0.342	0.316	0.313	0.349	0.322	0.322
City Controls	Y	Y	Y	Y	Y	Y

	IV: Number of Thermal Inversions					
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM2.5: 2006-2015)	-0.537*	-0.451	-0.953***			
	(0.284)	(0.299)	(0.335)			
Ln(PM2.5: 2011-2015)				-0.482*	-0.396	-0.896***
				(0.282)	(0.297)	(0.330)
Constant	0.895	-0.144	2.842	0.352	-0.672	2.195
	(2.327)	(2.443)	(2.744)	(2.249)	(2.364)	(2.633)
Observations	290	290	290	290	290	290
Adjusted R-squared	0.357	0.326	0.338	0.354	0.322	0.344
City Controls	Y	Y	Y	Y	Y	Y

Notes: Dependent variable is share of group that are immigrants. Independent variable is destination city PM 2.5. Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the number (top panel) and strength (bottom panel) of thermal inversions. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

A.1.2 The Endogenous Placement of Power Plants, and Predicting Thermal Inversions

In this section we test the identification assumptions underlying our two main instruments. First, we test the concern that plants that are located near cities are systematically built near poorer, less influential cities. Even though our instrument is the interaction between the three components of wind direction, distance and coal consumption, we may expect policy makers to take all three into account when placing plants near cities. In Table A7 we exclude any plants built within 200km of the city center (first two columns), and then within 400km of the center (last two columns). We still restrict the radius to be of 500km widths. Our results are similar to before, with a slight increase in precision.

Table A7: Different Distance Bins for Selection of Plants

	Dependent variable: Leave hukou city indicator			
	distance 200-700km		distance 400-900km	
	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0907**	0.171***	0.118***	0.190***
	(0.0437)	(0.0353)	(0.0376)	(0.0348)
Observations	384,221	77,825	384,221	77,825
Adjusted R-squared	0.027	0.070	0.023	0.066
City Controls	Y	Y	Y	Y
Demographics	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Even though policy makers may not have modeled the polluting potential of plants built in the past in the manner that we do, we may expect that newly built plants are subject to more scrutiny as the conversation about air quality in China has recently escalated. In Table A8 we exclude newly built plants in the IV, and instead include them in the 'control' group to be more conservative. We find similar patterns and magnitudes.

To test whether politicians avoid richer, influential cities when building new plants in a manner that simultaneously incorporates the interactions between wind direction, distance and coal consumption, we explore whether baseline city features predict newly built plants. In

Table A8: Excluding Newly Built Power Plants

	Plants > 5 yrs ago		Plants > 10 yrs ago		Plants > 15 yrs ago		Plants > 20 yrs ago	
	Low edu	High edu	Low edu	High edu	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0297	0.103**	0.0379	0.0914*	0.0389	0.0632	0.0609	0.0772*
	(0.0599)	(0.0434)	(0.0632)	(0.0467)	(0.0520)	(0.0438)	(0.0549)	(0.0435)
Observations	368,957	75,533	368,957	75,533	368,957	75,533	368,957	75,533
Adjusted R-squared	0.029	0.076	0.029	0.077	0.029	0.078	0.029	0.077
City Controls	Y	Y	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Regions affected by new plants included in sample (i.e. in the 'control' regions) so as to generate conservative estimates. Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A9: Baseline Economy and the Wind Direction IV

	The ratio of upwind plants		Wind direction and Coal plants IV	
Log Baseline Population	-0.009	-0.013	-93.886	-103.955
	(0.027)	(0.025)	(138.977)	(130.149)
Log Baseline GDP per capita	0.020	0.013	-165.748	-202.833
	(0.040)	(0.038)	(240.623)	(221.615)
Log Baseline Elec cons	-0.019		14.687	
	(0.021)		(97.437)	
Log Industrial Elec cons		-0.011		36.671
		(0.016)		(66.036)
Observations	276	276	277	277
Adjusted R-squared	0.006	0.004	0.492	0.493
City Controls	Y	Y	Y	Y

Notes: Dependent variables are for plants built post 2000, independent variables are measured in the year 2000. Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A9, we explore whether city-level characteristics in 2000 can predict (a) the ratio of upwind plants built after 2000, and (b) the IV based on plants built after 2000. We find no meaningful associations between these variables and possible predictors of a city's influence, like baseline populations, GDP, electricity consumption and industrial electricity consumption. In the following section, we also show that our results are robust to excluding big cities and major provincial capitals (Table A22).

Table A10: Lagged Pollution and Thermal Inversions

	Number of thermal inversion		Strength of thermal inversion	
Lagged Log(PM 2.5)	-1.603	-1.602	-3.632	-3.888
	(2.057)	(2.054)	(4.971)	(4.992)
Lag number of inversions		-0.0269		
		(0.0200)		
Lag strength of inversions				0.0478
				(0.0338)
Observations	5,392	5,392	5,392	5,392
Adjusted R-squared	0.249	0.250	0.225	0.226
Number of Cities	337	337	337	337
City Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. City level regressions for 337 cities over 16 years. Specifications include city and year fixed effects.

We also examine the claim that power plants may be built near centers that require more electricity. While it is hard to see why this should be correlated with wind direction, we still examine this claim seriously. In any case, most electricity is provided to the national or regional grid, and is less likely to be consumed nearby. In Table A9 we see that baseline electricity consumption or industrial electricity consumption are not strong predictors of the IV or the ratio of upwind plants.

Finally, we turn our attention to the thermal inversions IV. This IV has been used extensively by researchers in many different contexts (Arceo et al., 2016; Chen et al., 2017; Hicks et al., 2015; Jans et al., 2014), and as such, has been scrutinized thoroughly. Nonetheless, we examine whether lagged pollution levels can predict future levels of thermal inversions and the strength of inversions in Table A10. We fail to find any such meaningful associations. Furthermore, we also find that lagged inversions do not predict future inversions, suggesting that the levels of auto-correlation in inversions are low, and we may consider the data generating process underlying inversions to be close to random.

A.1.3 Different Panel Structures

We use the panel dimension of our data to further explore the relationship between out-migration and pollution. Since our instrument for wind direction and coal-fired power plants have little variation over time, we exclude that IV from this portion of the analysis, and instead concentrate on the OLS and thermal inversions IV. We note that the underlying variation in PM 2.5 for the panel and cross-section are likely to be similar. As Figure 2a shows, over time, the large increase in PM 2.5 was driven by an increase in the spread across cities. As such, the growth rate of PM 2.5 will reflect the cross-sectional variation in air quality in 2015.

In Table A11 we explore whether out-migration decisions depend on changes in pollution levels. We examine both changes in pollution between 2005-2015 and between 2010-2015. Our outcomes of out-migration probability are measured in 2015. Again, we consistently find effects on out-migration for the higher educated group of workers.

Table A11: Changes in Pollution and Out-migration

	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Δ Log (PM2.5)(2015-2010)	0.0226	0.00136	0.118*			
	(0.0678)	(0.0733)	(0.0624)			
Δ Log (PM2.5)(2015-2005)				-0.0367	-0.0503	0.0883**
				(0.0400)	(0.0432)	(0.0377)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.029	0.028	0.082	0.030	0.028	0.082
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: In the first three columns, the independent variable is the change in Log PM 2.5 between the average pollution in 2005-2010 and the average pollution in the 2010-15 period. The dependent variable of out-migration is measured in 2015. In the last three columns, the independent variables is the change in Log PM 2.5 between 2015 and 2005. Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

We then turn our attention to city-level panel regressions, where we can include city fixed effects and year fixed effects. At this point, we need to decide on what time period to measure

PM 2.5 over. It would be unlikely for migration decisions to depend on contemporaneous pollution shocks, as migration decisions are long-lasting. Indeed, in Table A12 we see that contemporaneous shocks in PM 2.5 have little detectable impacts on out-migration shares.

Table A12: Panel Fixed Effects (2005-15) with contemporaneous PM 2.5: City-level out-migration shares and PM 2.5

	2005 and 2015 OLS Out-migrant Share(%)			2005 and 2015 IV: strength of thermal inversion Out-migrant Share(%)		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM25)	-0.0824	0.00545	-0.0879*	-0.0922	-0.00224	-0.0899
	(0.0576)	(0.0125)	(0.0457)	(0.0941)	(0.0284)	(0.0668)
Observations	664	664	664	664	664	664
Number of Cities	332	332	332	332	332	332
Year Fixed Effects	Y	Y	Y	Y	Y	Y
City Fixed Effects	Y	Y	Y	Y	Y	Y

Notes: City level regressions where the migration outcome is measured over two years: 2005 and 2015. PM 2.5 levels are measured in those two years. First three columns are OLS, whereas the last three columns use the 'strength of thermal inversions' IV. Our sample consists of 332 cities. The share of outmigrants is the ratio of out-migrants to hukou city population. Standard errors clustered at the hukou city level are reported in parentheses. The weather controls include Average Annual Temperature, Average Annual Wind Speed, Average Annual Atmospheric Pressure, Average Annual Sunshine Duration, Average Annual Rainfall.

To that end, we turn our attention to specifications where PM 2.5 are measured over 5 years leading up to the migration decision. In Tables A13 - A15 we run city-level panel regressions with city and year fixed effects. Table A13 shows the OLS specification, whereas Table A14 is for the IV based on thermal inversions, and Table A15 is the IV based on the strength of thermal inversions. As we expect migration decisions to depend on persistent air quality, we use pollution measures aggregated up to the 5 preceding years in each specification.

Our decision to rely on the 5 years preceding the migration decision is somewhat arbitrary. In Tables A16 - A18 we examine different timing structures for our measures of PM 2.5. While we measure out-migration in 2015, we wish to understand how migration decisions depend on the cumulative PM 2.5 over different time intervals. While for the OLS specification, the results are less sensitive to different time periods, in the different IV specifications, we find that the longer the time period of PM 2.5 exposure, the larger is the response. As such, many shocks in a short time frame have a smaller response than the same number of shocks spread out over

Table A13: Panel Fixed Effects (2005-15): City-level out-migration shares and PM 2.5

	2005 and 2015 Out-migrant Share(%)			2005, 2010 and 2015 Out-migrant Share(%)		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM25) 5 year	0.304***	0.0820***	0.222***	0.283***	0.0578***	0.225***
	(0.0695)	(0.0106)	(0.0623)	(0.0266)	(0.00468)	(0.0234)
Observations	664	664	664	996	996	996
Number of Cities	332	332	332	332	332	332
Year Fixed Effects	Y	Y	Y	Y	Y	Y
City Fixed Effects	Y	Y	Y	Y	Y	Y

Notes: City level regressions where the migration outcome is measured over 2005-2015. PM 2.5 levels are measured as 5 year averages leading up these years. Our sample consists of 332 cities. The share of outmigrants is the ratio of out-migrants to hukou city population. Standard errors clustered at the hukou city level are reported in parentheses.

a longer time period.

Table A14: Panel IV Fixed Effects (2005-15): City-level out-migration shares and PM 2.5

	2005 and 2015 Out-migrant Share(%)			2005, 2010 and 2015 Out-migrant Share(%)		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM25) 5 year	0.572***	0.127***	0.445***	0.693***	0.167***	0.526***
	(0.170)	(0.0212)	(0.153)	(0.116)	(0.0189)	(0.102)
Observations	664	664	664	996	996	996
Number of Cities	332	332	332	332	332	332
Year Fixed Effects	Y	Y	Y	Y	Y	Y
City Fixed Effects	Y	Y	Y	Y	Y	Y

Notes: Migration outcome is measured over 2005-2015. Instrumental variables using the number of thermal inversions over the 5 preceding years. Our sample consists of 332 cities. The share of outmigrants is the ratio of out-migrants to hukou city population. Standard errors clustered at the hukou city level are reported in parentheses.

Table A15: Panel IV Fixed Effects (2005-15): City-level out-migration shares and PM 2.5

	2005 and 2015 Out-migrant Share(%)			2005, 2010 and 2015 Out-migrant Share(%)		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM25) 5 year	0.714**	0.151***	0.563**	-0.0398	-0.248	0.208
	(0.288)	(0.0380)	(0.255)	(0.816)	(0.183)	(0.709)
Observations	664	664	664	996	996	996
Number of Cities	332	332	332	332	332	332
Year Fixed Effects	Y	Y	Y	Y	Y	Y
City Fixed Effects	Y	Y	Y	Y	Y	Y

Notes: Migration outcome is measured over 2005-2015. Instrumental variables using the strength of thermal inversions over the 5 preceding years. Our sample consists of 332 cities. The share of outmigrants is the ratio of out-migrants to hukou city population. Standard errors clustered at the hukou city level are reported in parentheses.

Table A16: OLS: PM 2.5 Measured over Different Time Intervals

	Dependent variable: Leave hukou city indicator					
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
Ln(PM2.5: 1998-2015)	0.0523***	0.0513***	0.0612***			
	(0.0147)	(0.0160)	(0.0129)			
Ln(PM2.5: 2001-2015)				0.0505***	0.0488***	0.0614***
				(0.0147)	(0.0160)	(0.0129)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.032	0.030	0.084	0.032	0.030	0.084
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

	Dependent variable: Leave hukou city indicator					
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
Ln(PM2.5: 2006-2015)	0.0499***	0.0481***	0.0619***			
	(0.0144)	(0.0158)	(0.0126)			
Ln(PM2.5: 2011-2015)				0.0459***	0.0438***	0.0621***
				(0.0146)	(0.0158)	(0.0128)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.032	0.030	0.084	0.032	0.030	0.084
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: OLS regressions. Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A17: Number of Thermal Inversions IV: PM 2.5 Measured over Different Time Intervals

	Dependent variable: Leave hukou city indicator					
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
Ln(PM2.5: 1998-2015)	0.126***	0.124***	0.185***			
	(0.0389)	(0.0424)	(0.0415)			
Ln(PM2.5: 2001-2015)				0.122***	0.119***	0.178***
				(0.0374)	(0.0408)	(0.0396)
First stage F stat	22.56	22.56	22.56	23.32	23.32	23.32
Observations	462,046	384,221	77,825	462,046	384,221	77,825
Adjusted R-squared	0.024	0.023	0.066	0.024	0.023	0.067
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

	Dependent variable: Leave hukou city indicator					
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
Ln(PM2.5: 2006-2015)	0.112***	0.109***	0.168***			
	(0.0347)	(0.0378)	(0.0370)			
Ln(PM2.5: 2011-2015)				0.102***	0.0994***	0.159***
				(0.0332)	(0.0361)	(0.0354)
First stage F stat	23.52	23.52	23.52	23.42	23.42	23.42
Observations	462,046	384,221	77,825	462,046	384,221	77,825
Adjusted R-squared	0.026	0.025	0.069	0.026	0.025	0.071
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Instrumental variables using the number of thermal inversions measured over different time intervals. Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A18: Strength of Thermal Inversions IV: PM 2.5 Measured over Different Time Intervals

	Dependent variable: Leave hukou city indicator					
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM2.5: 1998-2015)	0.0832*	0.0854	0.176***			
	(0.0492)	(0.0541)	(0.0591)			
Ln(PM2.5: 2001-2015)				0.0794*	0.0812	0.169***
				(0.0470)	(0.0516)	(0.0560)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.031	0.029	0.074	0.031	0.029	0.075
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

	Dependent variable: Leave hukou city indicator					
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM2.5: 2006-2015)	0.0721	0.0733	0.160***			
	(0.0439)	(0.0480)	(0.0525)			
Ln(PM2.5: 2011-2015)				0.0598	0.0622	0.155***
				(0.0451)	(0.0498)	(0.0554)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.031	0.029	0.076	0.031	0.029	0.077
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

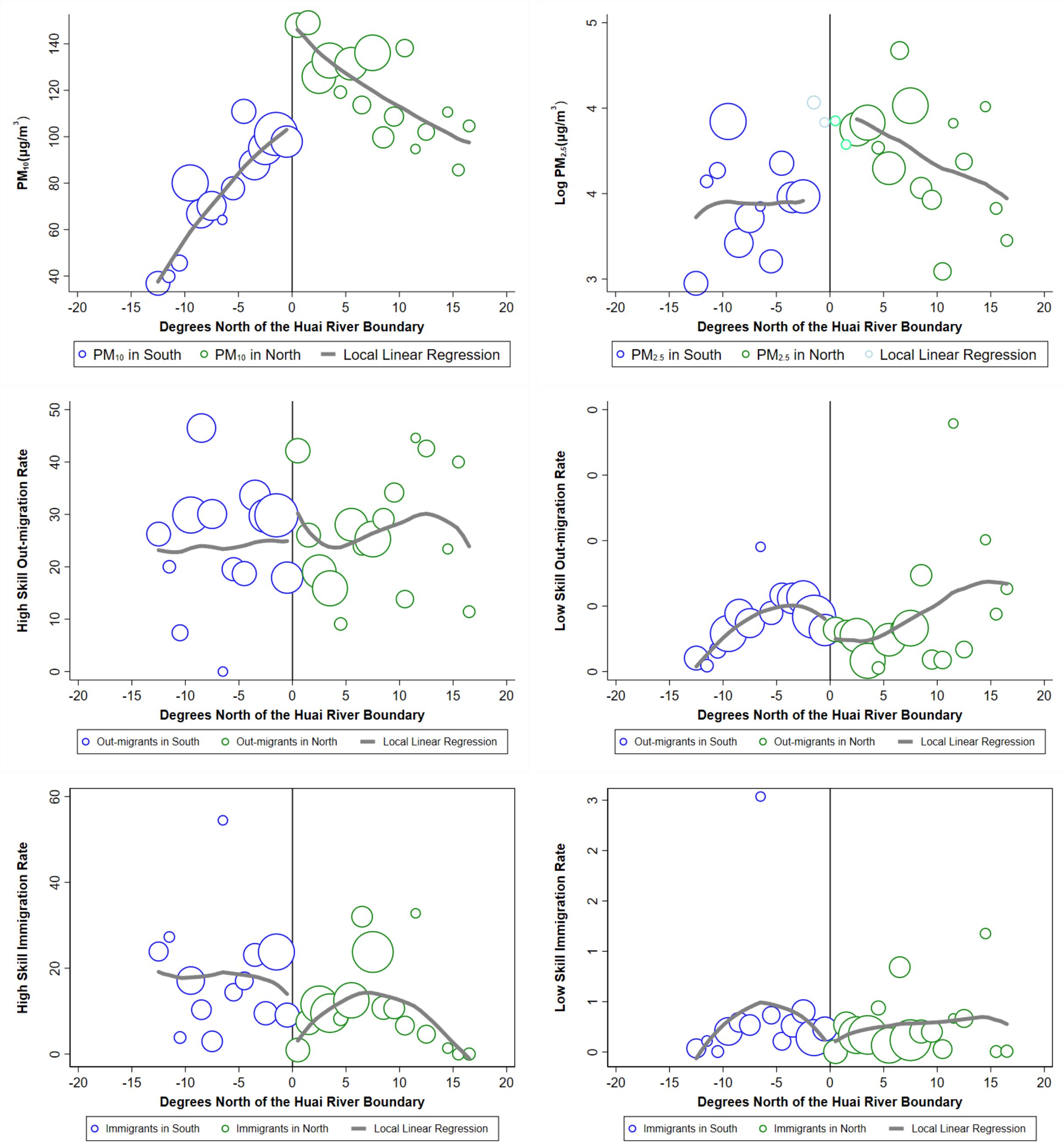
Notes: Instrumental variables using the number of strength of inversions measured over different time intervals. Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

A.1.4 The Huai River Regression Discontinuity

Between 1950-1980 China established free heating by providing free coal to residences and offices north of the Huai River. This policy had long lasting effects, as even today the heating systems are different between the northern and southern parts of the country. The north relies on coal boilers which releases a large amount of pollutants. Chen et al. (2013) examine the effects of this policy on life expectancy using a regression discontinuity design where they compare cities just north of the river to those just south of it.

Here, we leverage the same empirical setup to examine migration decisions. Figure 13 shows the RD graphs where the horizontal axis represents the distance between the city and the Huai river. In our top row we show the discontinuity in PM 10 and PM 2.5 levels. In the middle row, we look at out-migration decisions, and fail to find any meaningful changes in out-migration rates. In the bottom row, we look at in-migration rates, and find that only for the high-skill workforce, there is less in-migration in cities that have more pollution. This difference is statistically and economically meaningful in our RD regressions. We find no such differential response on in-migration for the low-skilled.

Figure 13: The Huai River RD



Notes: Top row show the discontinuity in PM 10 and PM 2.5 at the Huai River. Second row shows the out-migration by skill level. Bottom row shows the in-migration rate by skill-level. Bubble sizes are baseline city populations.

A.2 Alternative Model Specifications, Controls and Samples

In this section we examine different model specification, sample restrictions and control variables. Once again, our aim is to test the robustness of our estimates to various empirical concerns. In the interests of space, we restrict our results to the IV which produced the most conservative results – that of the wind direction and coal-fired power plants.

A.2.1 Different Sample Specifications

First, we use an alternative data source, the China Labor Dynamic Survey (CLDS), that allows us to define migration in a different manner. The 2014 survey asks a retrospective history of locations for individuals. We use this retrospective location history to create an individual-level panel between 2000 and 2014. We combine these panel with annual information on PM 2.5 at the city level. We define migration to be any change in location, whether or not the individual changed their *hukou* status. In these data, about 37% of low-skilled individuals, and 56% of high-skilled individuals changed their *hukou* location. Table A19 shows that once again, the response of high-skill migration to PM 2.5 is a lot stronger than that of low-skill workers. In the first three columns we look at all individuals, whereas in the last three columns we restrict it only to a sample of individuals who have ever changed city locations more than once.

Table A19: Individual-Level Panel (CLDS) Data

	Dependent Variable: Indicator for change city location					
	Whole Sample			Migrated>1 time		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Ln(PM25)	0.00155	-0.000249	0.0174***	0.00566	-0.00713	0.0674***
	(0.00130)	(0.00125)	(0.00622)	(0.00861)	(0.00927)	(0.0215)
Observations	239,471	210,418	29,052	38,756	30,958	7,796
Adjusted R-squared	0.064	0.073	0.028	0.031	0.041	0.000
Demographics	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y
City Fixed Effect	Y	Y	Y	Y	Y	Y
Individual Fixed Effect	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Data source is the China Labor Dynamic Survey 2014. Migration is defined as whether ever leaving a city to change location for at least one year. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Next, we look at different ways to slice the data. In Table A20, instead of splitting up the sample into low and high skilled, we split it up into three categories: high school graduates, those with some college education, and those with a college degree. A steep education gradient is apparent where the elasticity of migration with respect to PM 2.5 is higher for higher levels of education.

Table A20: Disaggregated Education Levels

	Dependent variable: Leave hukou city indicator		
	High school	Some College	College and above
Log(PM 2.5)	0.0318	0.0595*	0.131***
	(0.0527)	(0.0357)	(0.0435)
High school v some college diff p-val		0.16	
Some college v college diff p-val			0
Observations	384,221	40,656	37,169
Adjusted R-squared	0.029	0.079	0.086
City Controls	Y	Y	Y
Demographics	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

In Table A21 we split up the sample by rural and urban origin cities. We find that the elasticity for high-skill outmigration is larger in rural areas than it is in urban areas. In Table A23 we see that the effects are larger for younger workers, probably as that is the age group where location decisions are more salient.

Next, we examine whether big cities, high polluting cities, or major province capitals are driving our results. We do this by excluding such cities one at a time in Table A22. This may help allay concerns about the influence of major cities or province capitals in pollution policy, placement of plants, or being outliers in terms of pollutants and/or skilled jobs.

Finally, we test whether the wind direction and coal fired plants IV is influenced by the fact that plants may be located in coal producing regions (like Shanxi province). In as far as coal producing regions may also influence the underlying industrial structure, this may raise concerns about other unobservable associations with migration rates. In Table A24 we exclude these regions, and see that the results are more precisely estimated.

Table A21: Urban vs Rural Outmigration

	Dependent variable: Leave hukou city indicator			
	Urban hukou		Rural hukou	
	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0306	0.0816***	0.0341	0.211**
	(0.0370)	(0.0299)	(0.0757)	(0.0948)
High v Low edu diff p-val	0		0	
Observations	169,860	67,743	214,361	10,082
Adjusted R-squared	0.024	0.032	0.034	0.074
City Controls	Y	Y	Y	Y
Demographics	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A22: Without Big Cities, High Polluters and Major Province Capitals

	Exclude Beijing		Exclude Tianjin		Exclude Shijiazhuang	
	Low edu	High edu	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0196	0.0708*	0.103	0.0947**	0.0210	0.0657*
	(0.0569)	(0.0378)	(0.0750)	(0.0470)	(0.0573)	(0.0388)
Observations	366,760	72,655	363,518	72,261	367,439	75,096
Adjusted R-squared	0.029	0.085	0.031	0.083	0.029	0.085
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

	Exclude Shenyang		Exclude Zhengzhou		Exclude Wuhan	
	Low edu	High edu	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0194	0.0700*	0.0216	0.0763*	0.0195	0.0670*
	(0.0568)	(0.0381)	(0.0581)	(0.0396)	(0.0574)	(0.0394)
Observations	367,827	74,884	366,287	74,358	367,741	74,736
Adjusted R-squared	0.029	0.085	0.029	0.085	0.029	0.085
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A23: By Age Groups

	Dependent variable: Leave hukou city indicator					
	Age 25-34		Age 35-44		Age 45-54	
	Low edu	High edu	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0154	0.125**	0.0329	0.0574**	0.0399	0.0526***
	(0.0738)	(0.0562)	(0.0572)	(0.0270)	(0.0532)	(0.0177)
Observations	110,320	39,302	128,681	24,273	323,197	14,250
Adjusted R-squared	0.021	0.070	0.020	0.061	0.028	0.011
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A24: Without Coal Producing Regions

	Dependent variable: Leave hukou city indicator	
	Low edu	High edu
Log(PM 2.5)	0.0521	0.0999***
	(0.0452)	(0.0329)
High v Low edu diff p-val	0	
Observations	366,259	74,360
Adjusted R-squared	0.028	0.080
City Controls	Y	Y
Demographics	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

A.2.2 Different Model Specifications

Here we examine different model specifications for the wind-direction coal-fired plants IV. In Table A25 we try different versions of the instrument that exclude the distance or coal-consumption components from the instrument altogether. If we are concerned about the endogeneity of any one of these components, we can see that the results are robust to excluding them.

Table A25: Decomposing the Wind Direction IV

	IV:Excluding Distance			IV:Excluding Coal consumption		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.101**	0.0879*	0.154***	0.0622	0.0474	0.0886***
	(0.0458)	(0.0516)	(0.0365)	(0.0393)	(0.0455)	(0.0322)
Observations	462,046	384,221	77,825	458,363	380,978	77,385
Adjusted R-squared	0.027	0.027	0.073	0.031	0.030	0.082
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction and coal consumption at power plant (first two columns), and interaction between wind direction and distance to plant (next two columns). City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A26: Up vs Downwind Power plants

	IV: Both Up and Downwind		IV: Upwind only		IV: Ratio of Up/Downwind		
	Low edu	High edu	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0147	0.0662*	0.0752	0.136***	0.147*	0.125	0.256***
	(0.0543)	(0.0391)	(0.0634)	(0.0486)	(0.0875)	(0.0988)	(0.0988)
Observations	368,957	75,533	368,957	75,533	458,363	380,978	77,385
Adjusted R-squared	0.029	0.084	0.031	0.082	0.023	0.024	0.051
Additional Controls	None		Downwind IV		None		
F stat from upwind IV	27.79						
F stat from downwind IV	0.28						
City Controls	Y	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

In Table A26 we test the sensitivity of our results to excluding the downwind plants from the IV altogether, and we rely only on the up-wind plants for the IV. This explicitly ends up comparing two cities – one that has a similar plant downwind, and another that has a plant upwind. Our results are again a bit stronger and more precisely estimated if we restrict the IV to upwind plants only. In the last few columns we take this a step further and define our IV to actually be the ratio of up/down wind plants, and once again find strong precisely estimated effects on out-migration rates.

In Table A27 we create placebo instruments based on artificially changing the wind direction angle by first 90 degrees, and then 180 degrees. The first two columns show that as the angle is increased, the falsified instrument is less likely to predict PM 2.5. Whereas the last two columns show that the falsified IV is less likely to affect migration. If anything, the coefficient on migration is often negative when using the 180 degrees plus instrument, suggesting that perhaps cities that are upwind from plants may have more in-migration.

Table A27: Placebo Wind Directions

	Log(PM 2.5)		Dependent variable: Leave hukou city indicator			
			Full sample	High edu	Full sample	High edu
Coal IV Placebo	5.24e-07*		-7.86e-08	2.46e-08		
(wind direcetion plus 90 degrees)	(2.84e-07)		(5.99e-08)	(6.36e-08)		
Coal IV Placebo		7.24e-07			-1.90e-07**	-1.92e-08
(wind direcetion plus 180 degrees)		(5.01e-07)			(9.59e-08)	(1.04e-07)
Observations	285	285	444,490	75,533	444,490	75,533
R-squared	0.369	0.369	0.029	0.075	0.029	0.075
Demographics			Y	Y	Y	Y
City Controls	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction (falsified), distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Finally, in Table A28 we use the Air Quality Index (AQI) as our independent variable of interest. As sources of pollution affect not just PM 2.5 but also other pollutants, we may be picking up the combined impact of many pollutants. As such, we show that our results are robust to using the AQI as our independent variable of interest.

Table A28: Air Quality Index

	OLS			IV		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(mean AQI)	0.00644	0.00729	0.0398**	0.0473	0.0246	0.0970*
	(0.0193)	(0.0222)	(0.0194)	(0.0686)	(0.0760)	(0.0586)
Observations	441,943	366,717	75,226	441,943	366,717	75,226
First stage F stat					26.14	
Adjusted R-squared	0.029	0.028	0.081	0.029	0.028	0.080
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Independent variable is Log(mean Air Quality Index). Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

A.2.3 Additional Controls

In this section we include various sets of controls that may confound the association between the wind direction coal-fired plants IV and migration decisions. We may be concerned that nearby plants increase the amount of electricity supplied and thereby encourage industrial production. Even though most electricity flows into a regional grid, we examine this claim by controlling for electricity consumption in Table A29.

Also in Table A29, we control for other determinants of the demand for skilled work at baseline. We do this to check whether the potential for skilled work just so happens to be in places that are correlated with our IV. We find our estimates display similar patterns as before if we add controls for the teacher-student ratio, and the number of hospitals and doctors per capita.

In Table A30 we include controls for local economic production (first 3 columns) and natural amenities (last three columns). Our economic controls include baseline measures of populations, GDP per capita, the size of the services and manufacturing sector, and average wages. Our controls for natural amenities include rainfall, temperature, wind speed, sunshine duration and humidity. Our results are not meaningfully affected by these controls.

Importantly, in Table A31 we control for such weather and climatic factors when using thermal inversions as an instrument. Since thermal inversions may lead to changes in local

Table A29: Controlling for Electricity Consumption and Baseline Skill

	Dependent variable: Leave hukou city indicator						
Additional Controls:	Total Electricity Cons.		Industrial Electricity Cons.		Baseline Skill Indicators		
	Low edu	High edu	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0123	0.0754**	0.0151	0.0781**	0.0610	0.0474	0.0984**
	(0.0556)	(0.0340)	(0.0561)	(0.0345)	(0.0461)	(0.0503)	(0.0446)
Observations	357,333	73,943	357,333	73,943	342,638	281,395	61,243
Adjusted R-squared	0.031	0.089	0.031	0.089	0.041	0.040	0.089
City Controls	Y	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y	Y

Notes: First two columns control for electricity consumption and next control for industrial electricity consumption. Baseline skill controls at 2000 include the ratio of number of teachers to students, log of the number of books, hospitals, doctors per capita. Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

weather, it is important to account for such changes, and as such we control for a set of weather characteristics, and find our results to be more precisely estimated.

Table A30: Additional Controls for the Local Economy and Natural Amenities

	Additional Local Economy Controls			Additional Natural Amenities Controls		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0496	0.0384	0.0718*	0.0778*	0.0703	0.106***
	(0.0509)	(0.0597)	(0.0377)	(0.0413)	(0.0453)	(0.0410)
Observations	344,608	283,210	61,398	344,608	283,210	61,398
Adjusted R-squared	0.044	0.044	0.093	0.049	0.049	0.090
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Additional local economy controls at 2000 (first three columns) include the Log population, GDP per capita, industrial structure (the product value at sevice sector / manufacture sector), and average wage. Natural amenities (last three columns) include average rainfall, temperature, wind speed, sunshine duration and humidity. Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A31: Controlling for Weather and Natural Amenities

	IV: Number of Inversions			IV: Strength of Inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.100***	0.0968***	0.142***	0.0700*	0.0670*	0.121***
	(0.0316)	(0.0338)	(0.0306)	(0.0365)	(0.0383)	(0.0386)
Observations	367,962	304,016	63,946	367,962	304,016	63,946
Adjusted R-squared	0.045	0.044	0.086	0.047	0.047	0.088
City Controls	Y	Y	Y	Y	Y	Y
Natural Amenities	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Natural amenities controls include average rainfall, temperature, wind speed, sunshine duration and humidity. Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

B Appendix: Additional Tables and Figures

Table B1: Summary Statistics

Variable name	Description	Mean	Std. dev
Out migration	Indicator=1 if the person left his/her hukou city for more than 6 month, =0 otherwise	0.129	0.336
Female	Indicator=1 if the person is female	0.426	0.494
Age		40.097	11.529
Urban hukou	Indicator =1 if the person holds urban hukou, =0 otherwise	0.48	0.5
Married	Indicator =1 if the person is married	0.818	0.386
Secondary education	Indicator=1 if the highest degree is high school, =0 otherwise	0.181	0.385
Tertiary education	Indicator=1 if the highest degree is some college or above, =0 otherwise;	0.149	0.357
Pollution Levels	PM2.5	43.218	15.603
Population	Population	617.524	536.783
GDP. capita.	GDP per capita	48141.4	28581.71
Middle. Teacher	Middle school teacher per capita	0.004	0.001
Primary. Teacher	Primary school teacher per capita	0.004	0.001
Doctors	Doctors per capita	0.002	0.001
*Distance. Seaport**	Minimum distance to three large sea ports	622.075	476.05

We control for distance to the three largest seaports: Tianjin seaport, Shanghai seaport, and Shenzhen seaport. These seaports are located at the three major economic circles of China: Beijing-Tianjin-Hebei Metropolitan Region, The Yangtze River Delta, the Pearl River Delta.

Table B2: Examples of Hukou Restrictions

City	Beijing	Shanghai	Guangzhou	Shenzhen
Total hukou points needed	Varies	72	60	100
Education	Doctoral degree:37 point Master degree: 26 point Bachelor degree:15 point Some college:10.5 point	Doctoral degree:27 point Master degrees:24 point Bachelor degree:21 point	Above college: 60 point Some college:40 point High school: 40 point	Doctoral degree:100 point Master degrees:90 point Bachelor degree:80 point College:60 point
Skills		College English Test 6-8: 8 point College English Test 4: 7 point	Junior workers: 10 point Middle-level workers: 30 point High-level workers: 50 point	Junior workers: 20 point Middle-level workers: 40 point High-level workers: 70 point Senior technical worker: 100point Junior professional: 70 point Middle professional: 90 point Senior professional: 100 point

Table shows a few examples of *hukou* requirements for city workers.

Table B3: Preferences for Environmental Issues by Education Levels

Dependent Variable: The Environmental issue in China is Terrible	
Secondary education	0.133***
	(0.0122)
Tertiary education	0.176***
	(0.0145)
p-values	0.002
t-values	-3.17
The average value for primary educated	0.549
City Controls	Y
Demographics	Y
Residential city dummy	Y
N	24538
adj. R2	0.115

Data source: China Household Panel Survey 2016 (CFPS). Secondary education denotes workers whose highest degree is high school. Tertiary education denotes workers whose highest degree is some college or above. In the CFPS 2016, there is a survey question: In your opinion, how terrible the environment issue is in China. (0=totally not terrible; 2,···,10=very terrible). Based on this question, we define environmental attitude dummy: D=1, if the answer is 6-10; =0, if the answer is 0-5. P-value: the p-values of test of Tertiary education =Secondary education; t-value: t-values of test of Tertiary education =Secondary education. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table B4: Concerns and Actions Taken on Environmental Issues

Panel A: Concerns on Environmental Issues

	Discuss environmental issues	Donation for environment protection	Concern over environmental issue
Secondary education	0.134***	0.0585***	0.120***
	(0.0171)	(0.0115)	(0.0154)
Tertiary education	0.231***	0.151***	0.174***
	(0.0188)	(0.0153)	(0.0189)
P-value	0.000	0.000	0.001
t-value	-6.01	-5.71	-3.56
Baseline average	0.392	0.106	0.402
City indicator	Y	Y	Y
City Controls	Y	Y	Y
Demographics	Y	Y	Y
N	11147	11147	11147
adj. R2	0.190	0.171	0.192

Panel B: Actions Taken on Environmental Issues

	Appeal on Environmental issue	Government environmental activity	Non-government environmental activity
Secondary education	0.0255***	0.119***	0.0690***
	(0.00845)	(0.0140)	(0.0118)
Tertiary education	0.0574***	0.246***	0.156***
	(0.0120)	(0.0169)	(0.0144)
P-value	0.010	0.000	0.000
t-value	-2.62	-8.21	-6.31
Baseline average	0.0597	0.135	0.100
City indicator	Y	Y	Y
City Controls	Y	Y	Y
Demographics	Y	Y	Y
N	11147	11147	11147
adj. R2	0.176	0.186	0.194

Data source: Chinese General Social Survey (CGSS). Secondary education denotes workers whose highest degree is high school. Tertiary education denotes workers whose highest degree is some college or above. In the CGSS, there is a survey question: whether you participate in the following activity. 1=never, 2=occasionally; 3=often. We define an indicator: D=1 if the answer=2,3; D=0 if the answer=1. P-value: the p-values of test of Tertiary education =Secondary education; t-value: t-values of test of Tertiary education =Secondary education. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table B5: Skill Ratio and House Prices

	OLS		IV: Higher ed expansion	
	Dependent Variables:		Log Rental Price	
$Log(L_s/L_u)$	0.905***	1.114***	1.164***	1.238***
	(0.0657)	(0.00154)	(0.0725)	(0.00147)
Weights	No	Population 2000	No	Population 2000
Observations	300	300	300	300
R-squared	0.452	0.612	0.424	0.607
City Controls	Y	Y	Y	Y

Rental price data calculated using household level rents in the China mini census of 2005. Instrumental Variable estimates leverage the expansion of higher education institutions between 2001. The first stage uses the share of college graduates by city. Standard errors clustered at the hukou city level are reported in parentheses. City controls include the log distance to Shanghai seaport, to Tianjin seaport, to Shenzhen seaport and to the nearest big city.

C Appendix: Additional Model Derivations

In this appendix we derive the labor supply curve from the worker utility function.

$$V_{sjdk} = \epsilon_{jd} w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d exp^{-H_{sd}} \tag{7}$$

Workers will pick the destination with the highest value of $V_{sjdk} = \widetilde{w_{sd}} \epsilon_{jd} exp^{-H_{sd}}$, where we define $\widetilde{w_{sd}} \equiv w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d$ to be a composite of wages and amenities. The probability that someone from origin o picks destination 1 is given by:

$$\begin{aligned}
\pi_{1os} &= Pr\left[\widetilde{w_{1s}} \epsilon_1 exp^{-H_{s1}} > \widetilde{w_{d's}} \epsilon_{d'} exp^{-H_{sd'}}\right] \quad \forall d' \neq 1 \\
&= Pr\left[\epsilon_{d'} < \frac{\widetilde{w_{1s}} exp^{-H_{s1}} \epsilon_1}{\widetilde{w_{d's}} exp^{-H_{sd'}}}\right] \quad \forall d' \neq 1 \\
&= \int \frac{dF}{d\epsilon_1} \left(\epsilon_1, \omega_1 \epsilon_1, ..., \omega_D \epsilon_D\right) d\epsilon_1
\end{aligned} \tag{24}$$

where we define $\omega_d \equiv \frac{\widetilde{w_{1s}} exp^{-H_{s1}}}{\widetilde{w_{d's}} exp^{-H_{sd'}}}$. We assume that the abilities are distributed with the following Frechet distribution:

$$F(\epsilon_1, ..., \epsilon_D) = exp\left\{-\left[\sum_{d=1}^{D} \epsilon_d^{-\eta}\right]\right\} \tag{25}$$

So the derivative of the CDF is given by:

$$\frac{dF}{d\epsilon} = \eta \epsilon^{-\eta-1} exp\left\{-\left[\sum_{d=1}^{D} \epsilon_d^{-\eta}\right]\right\} \tag{26}$$

This derivative evaluated at $(\epsilon_1, \omega_1 \epsilon_1, ..., \omega_D \epsilon_D)$, allows us to determine the probability of choosing destination 1, given by π_{1os}:

$$\begin{aligned}
\pi_{1os} &= \int \eta \epsilon^{-\eta-1} exp\left\{-\left[\sum_{d=1}^{D}(\omega_d \epsilon)^{-\eta}\right]\right\} d\epsilon \\
&= \frac{1}{\sum_{d=1}^{D}\omega_d^{-\eta}} \int \left(\sum_{d=1}^{D}\omega_d^{-\eta}\right)\epsilon^{-\eta-1} exp\left\{-\left[\epsilon^{-\eta-1}\left(\sum_{d=1}^{D}\omega_d^{-\eta}\right)\right]\right\} d\epsilon \\
&= \frac{1}{\sum_{d=1}^{D}\omega_d^{-\eta}} \int dF(\epsilon) \\
&= \frac{1}{\sum_{d=1}^{D}\omega_d^{-\eta}} .1 \\
&= \frac{\left(\widetilde{w_{1s}} exp^{-H_{s1}}\right)^{\eta}}{\sum_{d=1}^{D}\left(\widetilde{w_{ds}} exp^{-H_{sd}}\right)^{\eta}} \qquad (27)
\end{aligned}$$

The third line comes from the properties of the Frechet distribution, where we know that the term in the integral of the second line is simply the PDF with a shape parameter η, and a scale parameter $\sum_{d=1}^{D}\omega_d^{-\eta}$. Expanding on the definitions for $\widetilde{w_{ds}}$, and scaling up the probability by the size of the skilled workforce P_s in the country, we have:

$$L_{sd} = \frac{\left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d exp^{-H_{sd}}\right]^{\eta} P_s}{\sum_{d'}\left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{d'} exp^{-H_{sd'}}\right)^{\eta}} \qquad (28)$$

The Frechet assumptions also allow us to measure aggregate welfare. Using Equation 7, we can integrate over the the location preference ϵ_{jd}, conditional on choosing a destination.

$$\begin{aligned}
E[V_{sjdk}|d] &= \left(\widetilde{w_{ds}} exp^{-H_{sd}}\right) E[\epsilon_{jd}|d] \\
&= \left(\widetilde{w_{ds}} exp^{-H_{sd}}\right) \pi_{dos}^{-\frac{1}{\eta}} \Gamma\left(1-\frac{1}{\eta}\right) \\
&= \left(\sum_{d'}\left(\frac{w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{d'} exp^{-H_{sd'}}}{P_s}\right)^{\eta}\right)^{\frac{1}{\eta}} \Gamma\left(1-\frac{1}{\eta}\right) , \qquad (29)
\end{aligned}$$

where Γ is the gamma function, and is constant across cities and skill groups. Average city utility may depend on *hukou* costs. For instance, if a high-amenity city has a very restrictive *hukou* policy it may have a high average utility as not enough people can enter and lower wages and raise house prices. We define average utility for those in city d to be:

$$\overline{V_{sd}} = exp^{H_{sd}} \left(\sum_{d'} \left(\frac{w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{d'} exp^{-H_{sd'}}}{P_s} \right)^{\eta} \right)^{\frac{1}{\eta}} \equiv exp^{H_{sd}} \zeta_s, \quad (30)$$

The equation shows that the average utility depends on the average option value migrating to any other city, and the utility earned there. This average is scaled by the Frechet shape parameter η as it captures the dispersion in tastes across locations. Finally, the average utility of those in city d may be increasing in *hukou* restrictions, as they prevent immigrant workers from entering desirable cities, lowering wages and raising pollution levels. Yet, the utility of those in city d is a decreasing function of *hukou* restrictions in all other cities, as the option value of moving to those cities fall. We can therefore, rewrite the average utility as a function of *hukou* restrictions, and the labor supply as a function of utility in the manner described in the text:

$$L_{sd} = \frac{\left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d \right]^{\eta}}{\overline{V_{sd}}^{\eta}} \quad \& \quad Log\ \overline{V_{sd}} = H_{sd} + Log\ \zeta_s \quad (9')$$

Under a different set of assumptions, we will derive a similar relationship. For instance, if the *hukou* restrictions only affect potential immigrants, and not natives, then we first need to solve for native and immigrant populations separately. The number of native n individuals choosing to remain in city d is given by:

$$L_{sdn} = \frac{\left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d \right]^{\eta} \ P_{sd}}{\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{d'} exp^{-H_{sd'}} \right)^{\eta}} \ , \quad (31)$$

where P_{sd} is the population of city d. For migrant workers, the cost of entering city d will also depend on the *hukou* restrictions.

$$L_{sdm} = \left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d exp^{-H_{sd}} \right]^{\eta} \sum_{h \neq d} \frac{P_{sh}}{\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{d'} exp^{-H_{sd'}} \right)^{\eta}} \quad (32)$$

Such a setup generates an upward sloping labor supply curve, where the total supply of workers in skill s is given by:

$$L_{sd} = \frac{\left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d \right]^{\eta}}{\overline{V_{sd}}^{\eta}} \quad (9)$$

where $\frac{1}{\overline{V_{sd}}^{\eta}}$ captures the population weighted disutility of being in city d.

$$\frac{1}{\overline{V_{sd}}^{\eta}} \equiv \left[exp^{-H_{sd}\eta} \sum_{h \neq d} \frac{P_{sh}}{\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{d'} exp^{-H_{sd'}} \right)^{\eta}} + \frac{P_{sd}}{\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{d'} exp^{-H_{sd'}} \right)^{\eta}} \right] \tag{33}$$

We can rewrite this relationship, in a manner similar to before, as in Equation 9':

$$Log\ \overline{V_{sd}} = -Log\ \left(\frac{exp^{-H_{sd}\eta} P_{-ds} + P_{sd}}{P_s} \right) + Log\ \zeta_s \tag{34}$$

Once again, $\overline{V_{sd}}$ is an increasing function of *hukou* restrictions, but this time weighted by the fraction of skilled workers that live outside the city. The utility for individuals in all other cities is a decreasing function of city d's *hukou* restrictions, as it lowers the option value for potential migrants of moving to city d. If cities are small relative to the total population of the country (for instance, only about 1.5% of China's population lives in Beijing), then Equation 34 approaches the result in Equation 9'.