The Productivity Consequences of Pollution-Induced Migration in China

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

October, 2020

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 individual and city fixed-effects, and when instrumenting for pollution using distant power-plants upwind of cities, or 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 larger than the direct health benefits of clean air. *Hukou* migration-policy restrictions on mobility exacerbates the losses in productivity and welfare.

Keywords: Internal migration, air pollution, spatial productivity gaps
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-NUS University, ran.song@yale.edu
We thank seminar participants at UCSD, UW (Seattle), NBER Summer Institute (Energy and Enivornmental Economics, and the Urban Economics), Stanford SITE Migration Conference, U Richmond, Cities and Development Conference (Johns Hopkins), Ortygia Business School, Georgetown University, Barcelona UAB, the World Bank, Harvard, CUHK, HKUST (China Summer Institute), Nankai University, 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, as 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. Particulate matter pollution exceeds WHO air quality guidelines for 96% of Chinese cities in 2015, and on average is 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 migrate between cities.

We analyze whether workers relocate in response to variation in air quality across Chinese cities, and then quantify 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 the skilled 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 distant 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 individual-level 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 (Giles et al., 2019). 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 the skilled and unskilled 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 labor supply curve, and produce a valid estimate of skill-specific labor supply responses to changes in wages.[1]

Our quantification exercise is enriched by the fact 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 resorting 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 Kunming would bridge the between-city skilled wage gap by one-third.[3] The fact that pollution explains a meaningful portion of the productivity

[1]Pierce and Schott (2016) and Autor et al. (2013) use "import" shock to document effects on the United States, and we recognize that 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 an instrument for export-induced growth across Chinese cities.

[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*

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. This phenomenon is not relegate to China: When 9,000 Delhi residents were asked about their plans to deal with pollution, their single-most common response was "relocate" (Kapur, 2019).[4]

To quantify the productivity loss from pollution for a given city, we perform a set of counterfactuals for Beijing. We first halve the amount of overall pollution in Beijing (a policy that mimics pollution caps for the city), or halve simply the non-production, non-human component of pollution (say, by relocating nearby coal-fired plants). In each scenario, GDP per worker rises by more than 11%, mostly as a consequence of the relocation of workers to Beijing. Unskilled wages in Beijing rise by as much as 15% as more (complementary) skilled workers enter the city. This wage effect is once again largely driven by the relocation of workers rather than the direct health-driven productivity changes.

To understand the overall consequences of the *location* of pollution within the country, we 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 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.[5] This exercise increases aggregate GDP in China by about 4.7%. The relocation of workers is a major driver of these effects, larger than 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. It's important to study this relationship because increased pollution and migration have 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).

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 C2). Without the exemption, the system imposes a

pollution exodus: Firms struggle to woo top talent" https://phys.org/news/2019-03-asia-pollution-exodus-firms-struggle.html

[4]Indeed, 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). Recent reports following wildfires in California suggest that this may not be solely a developing world phenomenon (Lustgarten, 2020).

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

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 raise the welfare of skilled workers by as much as 26.6% and unskilled workers by 1.63%.

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 11.6%.[6]

Other research documents Chinese households' willingness to pay to avoid pollution using variation in wages and housing prices (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.[7] 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.[8] In contrast, the first part of our paper uses the restricted-access 2015 One-percent Population Census of China on actual migration decisions (after information about pollution was widely disseminated) and a longitudinal panel data which track individual migration over time from 2006 to 2014 (before and after information about pollution was widely disseminated) to explore the relationship between pollution and internal migration.

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 8. Section 9 concludes.

[6]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, posing a disproportionate burden on the poor. As industrialization in many developing countries worsens environmental quality, our results suggest that high migration costs exacerbate the welfare and productivity losses from pollution.

[7]See, for instance,"*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

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

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, gender, education level, employment status, occupation, *hukou* location, 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.[9] *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).[10] *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 longitudinal panel using the China Labor-force 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 2006 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 net number of skilled and unskilled workers in each city. We 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

[9] 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.

[10] In China, *hukou* plays a critical role as an internal-passport which determines one's entitlements to pursue many activities and eligibility for state-provided goods and services in a specific place. The migrants who do not hold a local *hukou* have limited or no access to many government-provided benefits, including public education for children and medical care.

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 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. 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 compare 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).[11] 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 by MEP to measure city-level air quality, which 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.

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

[11] 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/.

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) to the city-industry level using the 2-digit industrial classification for manufacturing industries.[12] 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 Regional 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.

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 employ these variables as controls. Appendix Table C1 reports the summary statistics and a full description of the key variables used in the analysis.

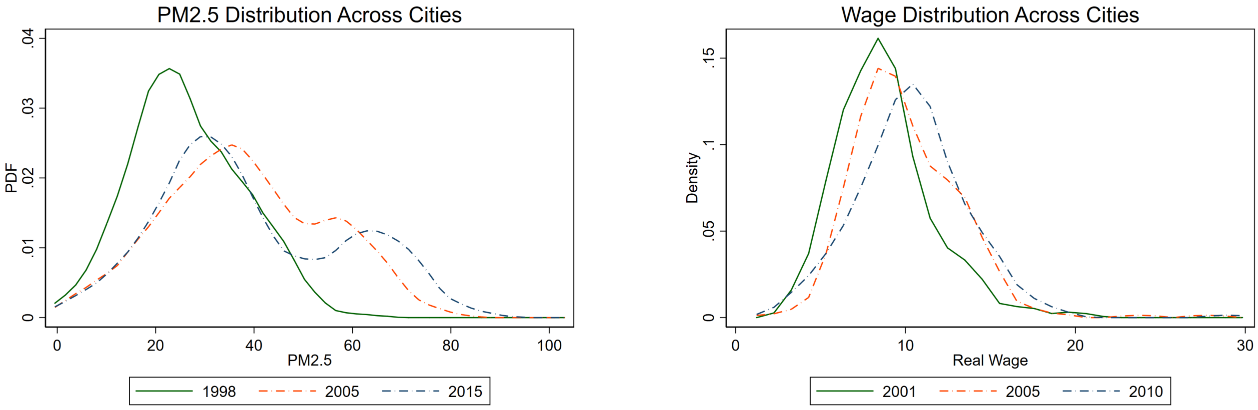
3 Descriptive Patterns of Pollution and Migration

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.

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

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 2a illustrates the time trend of annual PM2.5 concentrations in Chinese cities since 1998. The mean concentration exceeds WHO air quality guideline every year.[13] 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.

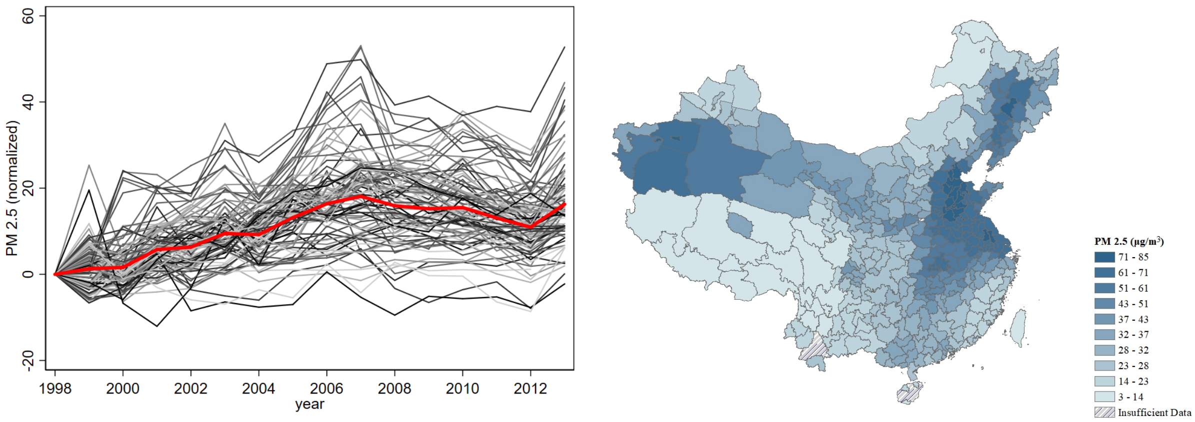
Next, we explore the spatial variation in air quality. Figure 2b illustrates the annual average satellite PM2.5 concentration data for 2015. Air quality is unevenly distributed across China. The coastal areas of north-east and eastern China experience the most severe air pollution. Manufacturing industries are concentrated in the east. The north-east further suffers from

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

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



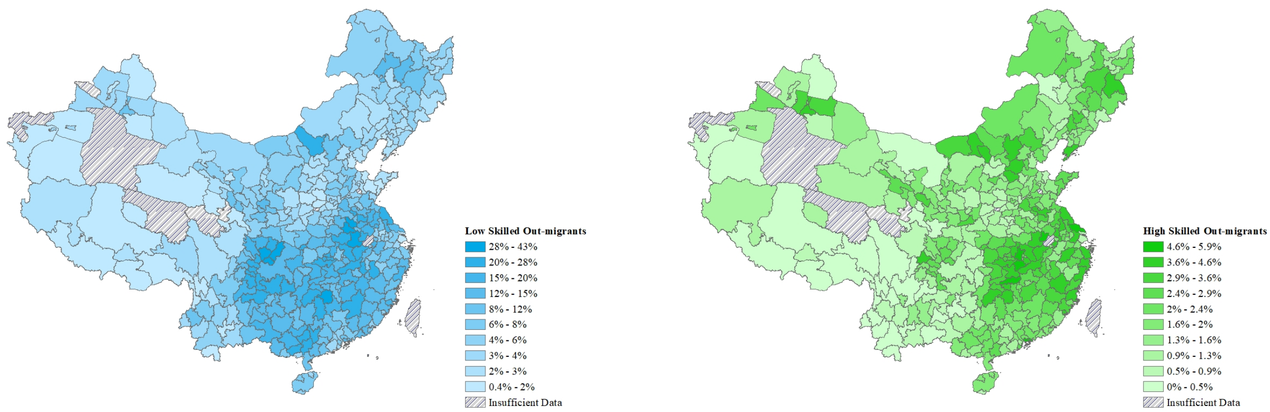
Notes: The spatial and temporal distribution of PM 2.5 using the Global Annual PM 2.5 Grids. The map shows the geographic spread in 2015. Figure 2a shows the increase in PM 2.5 over time for the 100 largest cities in China, where the 1998 value of PM 2.5 is normalized to be 0.

coal-burning due to heating needs, which exacerbates pollution even relative to high economic growth areas of the south.[14]

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.

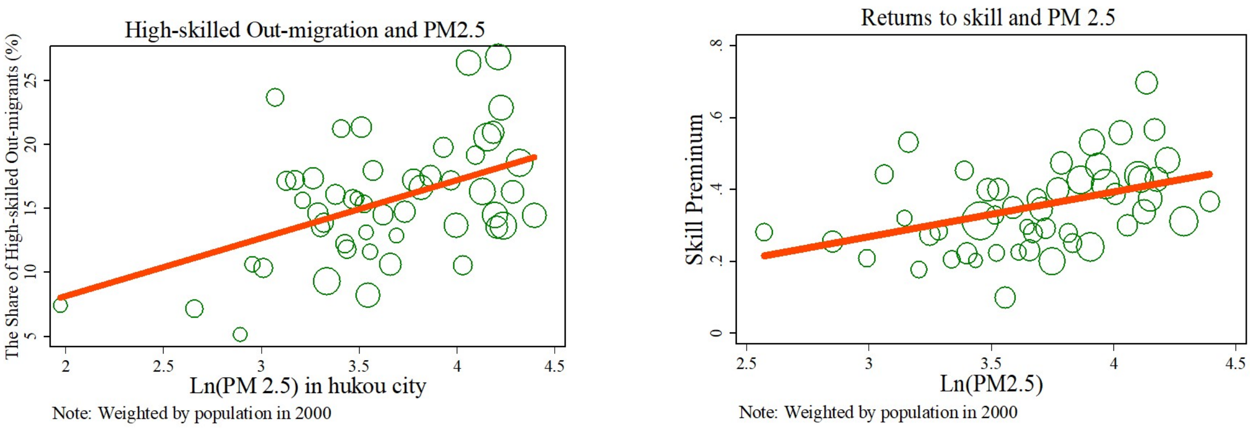
Correspondingly, we examine the geographic patterns of emigration of low-skill (Figure 3a) and high-skill (Figure 3b) migrants. Low-skill emigration rates are very high in the south of China, while high-skill out-migrants are comparatively more populous in the north-east and the east. Recall from Figure 2b that pollution is also relatively more concentrated in the north-east

[14]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.

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, suggesting 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 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 some college (or above)-educated out-migrants from the city-level hukou population. Returns to Skill denotes the return to some college or above education. Each bubble is a city, and the bubble size is weighted by the population in 2000.

Finally, Figure 4b examines the association between pollution and the wage returns to skills that emerges in each city.[15] 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 so their marginal product may be relatively higher), and this sorting could be a mechanism that raises aggregate productivity.

[15]These returns are consistent with recent estimates from other work, such as Giles et al. (2019).

4 Identifying the Causal Effect of Pollution on Migration

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

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

where M_{id} is an indicator for whether or not individual i left city d, and $\mathbf{X}$ are controls.

Before 2012, information about local PM2.5 concentration was not available in most Chinese cities. The Chinese government started to release PM2.5 data in 2012, and the unexpected data disclosure affected the avoidance behavior of Chinese citizens (Jia et al., 2019).[16] As such, in our preferred specification, we use the most recent census data in China–the 2015 One-Percent Population Census. In our robustness checks, we employ an individual longitudinal panel data, which allow us to account for city-level and individual-level fixed effects.

The OLS approach may produce 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 and a range of local disamenities, as polluted areas may have high manufacturing-based economies and poor government regulation. To get around these issues, we use a few different identification strategies to isolate the effect of pollution not driven by economic activity and other disamenities. In addition to OLS and panel fixed-effects models, we discuss in detail two different instrumental variables strategies. Appendix A includes 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 the economic value of 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:

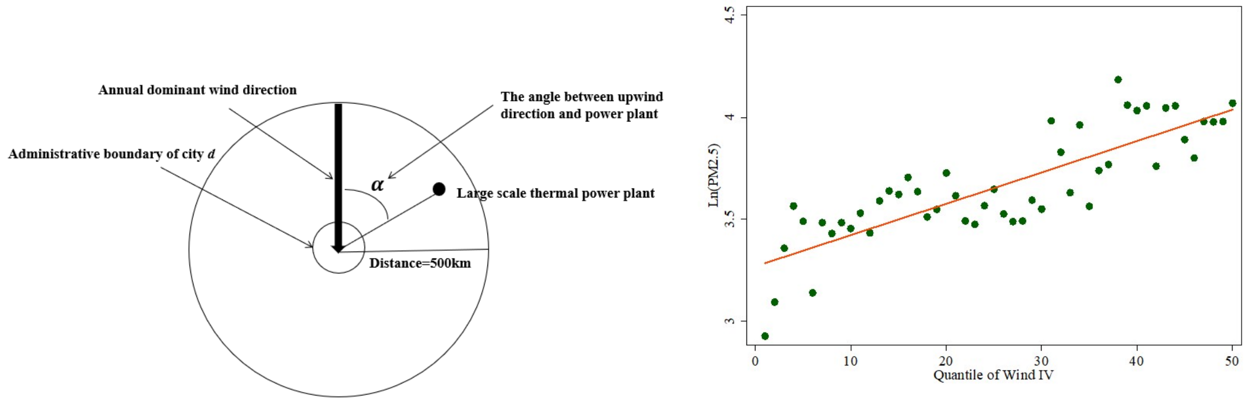
$$Log(PM2.5)_d = \gamma_0 + \gamma_1 \sum_{p}^{P} \left(\frac{1}{\alpha_p + 1}\right) \left(\frac{1}{dist_{pd}}\right) C_p + \varepsilon_j \ , \tag{2}$$

where α_p denotes the angle between the annual prevailing wind direction of city d and the large-scale power plant p, $dist_{pd}$ is the distance between the plant p and city d, 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* and within a 500km radius from the city center. Figure 5a explains

[16] Appendix B shows that the disclosure of PM2.5 data has an important impact on internal migration and air purifier purchase in China.

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.

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



(a) Depiction of instrument

(b) First-stage relationship

Notes: In the left panel, the thick arrow represents the annual dominant wind direction of city d. The dark dot represents a large-scale thermal power plant located outside city d and within 500 km from the city. The angle α denotes the angle between the annual prevailing wind direction of city d and the large-scale power plant. The large-scale thermal power plants are defined as the thermal power plants whose installed-capacities are larger than 1 million KW. In the right panel, cities are grouped into one hundred groups according to the quantile of the wind direction IV measure. The y-axis denotes the mean value of PM2.5 in each quantile and x-axis denotes the mean value of wind direction IV in each quantile.

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. Second, the large-scale thermal power plants supply electricity to vast areas of China; many 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 placed away from important cities, we show robustness to excluding power plants in a 200km radius away from cities. Among other specification tests, we show the robustness to excluding richer or capital cities, coal producing regions, using other outcome measures of air quality, and additional controlling for electrification, demographics and industrialization. The concerns around pollution are relatively recent, and we show robustness to using only old power plants, such as those built more than twenty years prior to our data. We also conduct numerous falsification tests showing that baseline city characteristics do not predict the future placement of plants, and tests with placebo wind directions indicate that 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 a meteorological phenomenon where the above-ground temperature is abnormally higher than the ground temperature, trapping pollutants. It is a strong predictor of poor air quality. We create two measures of inversions TI_d in city d. First, we count the number of thermal inversions in each year t. Next, we measure the annual mean strength of these inversions. We use this measure in both the cross-section and panel form. Our panel specifications include year fixed effect τ_t and city fixed effects δ_d:

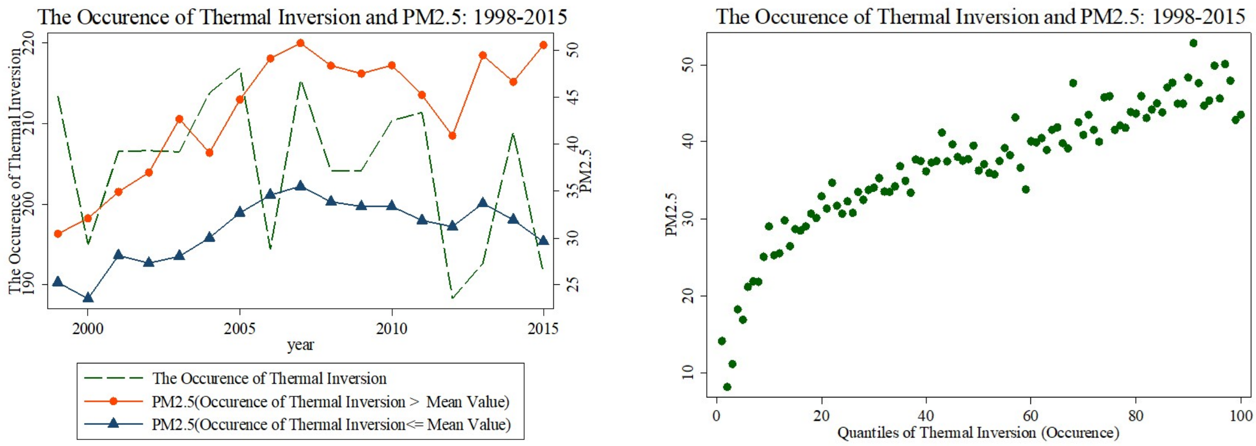
$$Log(PM2.5)_{dt} = \gamma_0 + \gamma_1 TI_{dt} + \delta_d + \tau_t + \varepsilon_{dt} \quad (3)$$

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 relatively more inversions over the decade saw 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.

In our specifications we control for time-varying natural amenities like sunshine and weather, show variants of our measures of inversions, and show how past pollution levels do not predict future inversions.

Together, these two instruments capture the variation in air quality due to either wind direction or meteorological phenomenon, and are less likely to be directly related to local economic activity. In specification checks in Appendix A, we compare different instruments and their combinations, lagged and accumulated panel structures, relevant controls, and exclude important cities. Finally, in Appendix A.1.3 we introduce an alternative source of variation – the Huai-river Regression Discontinuity (RD) to support our results (Chen et al., 2013).

Figure 6: Thermal Inversions and Air Quality



Notes: In the left panel, we divide cities into the two groups based on whether or not they lie above the average annual occurrence of thermal inversions. The red line represents the mean value of PM2.5 in cities where the occurrence of the thermal inversions are above average. The violet line represents the mean value of PM2.5 in cities where the occurrence of the thermal inversions are below average. The green-dash line presents the average annual occurrence of thermal inversions. In the right panel, cities are grouped into one hundred groups according to the quantile of the occurrence of thermal inversions. The y-axis denotes the mean value of PM2.5 in each quantile and x-axis denotes the mean occurrence of thermal inversions in each quantile.

5 Empirical Results

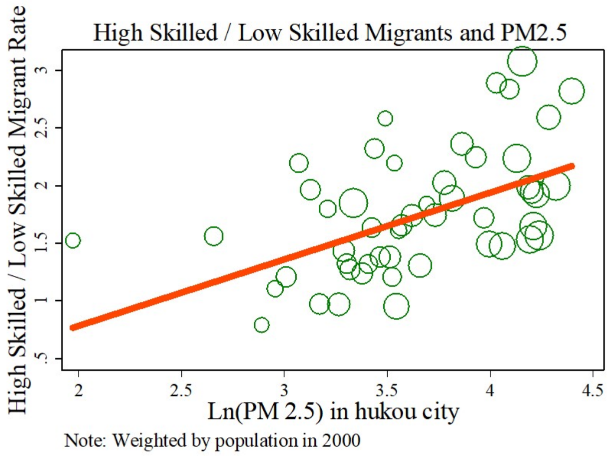
Here, we discuss our main empirical relationships between migration and pollution. We relegate most of the extensive discussion on our empirical results to Appendix A, which includes the long set of specification tests, falsification tests, sample restrictions, different panel-data structures, and different 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. On the vertical axis we plot the ratio between the out-migration rates for high-skilled and low-skill workers. This ratio increases with higher PM 2.5 levels in source cities, suggesting that the out-migration for high-skilled workers is stronger than that of low-skilled workers.

In Table 1 we examine the 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 some college degree or above and those without. The OLS estimates in Panel A suggest that air pollution leads to the out-migration of workers, and the impacts are stronger for those with higher education. Yet, air pollution is likely to be correlated with economic activity and other local disamenities, biasing naive OLS estimates. In the top panel, we employ our first instrument based on wind direction and distant coal-fired power plants (Freeman, Liang, Song, and Timmins, 2019) to deal with the endogeneity issue.The differential impact by skill level increases in magnitude when using our instruments to address the

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.

endogeneity. Our IV results suggest that a 10% increase in PM 2.5 raises out-migration rates by 0.543 percentage points, with the effect being meaningfully larger for those with higher education attainment (1.02 percentage points) than those without (0.302 percentage points).

In the lower panel of Table 1, we study how variation in PM 2.5 from our second instrument of thermal inversions affects migration tendency. We include distances to three large seaports to account for the spatial distribution of economic development in China. In the first three columns, we employ the annual occurrence of thermal inversions as the instrumental variable. The results show that the implications of air pollution on emigration are more pronounced for high-skilled workers in comparison with those with lower skills. A 10% increase in PM 2.5 leads to a 1.79 percentage point increase in out-migration rates for those with high education attainment, but only 1.08 for those without. In the last three column, we leverage the variation coming from the annual strength of thermal inversions. The results remain similar.

In Appendix A.1.1 we study our results across different instrumental variable strategies. In Table A1 we show the first stage of the different instruments, all of which display a strong and robust first-stage relationship between our instruments and atmospheric PM 2.5. Since thermal inversions may be affected by weather conditions, we account for local weather conditions in Tables A3, and find a similar empirical pattern. In Table A4 we try combinations of the different instruments.

Tables A5-A6 show that the in-migration decisions also depend on PM 2.5 concentration in destination cities. Together these results suggest that workers leave polluted areas and seek out less polluted cities. Importantly, this response to pollution is stronger among higher educated workers. 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.

Table 1: Pollution and Out-Migration

	Dependent variable: Leave hukou city indicator					
	OLS Regression			Wind+Coal IV		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0497***	0.0471***	0.0571***	0.0543	0.0302	0.105**
	(0.0146)	(0.0156)	(0.0150)	(0.0502)	(0.0596)	(0.0433)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.030	0.029	0.078	0.030	0.029	0.076
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

	Dependent variable: Leave hukou city indicator					
Instruments:	Number of inversions IV			Strength of inversions IV		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.113***	0.108***	0.179***	0.0790*	0.0733	0.189***
	(0.0373)	(0.0412)	(0.0480)	(0.0424)	(0.0465)	(0.0671)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.027	0.026	0.066	0.030	0.029	0.064
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 in the top panel uses the interaction between wind direction, distance to coal plant, and coal consumption at power plant. Instrumental variables specifications in the bottom panel uses thermal inversions. 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.

5.2 Robustness and Heterogeneity

In Appendix A.1.2 we examine threats to identification for our various sources of variation. We seriously evaluate the claim that the location of power plants may depend on the simultaneous combination of wind direction, distance to cities, and the amount of coal consumed. First, power plants may be systematically built near poorer, less influential cities, and so the instrument may be correlated with unobservable characteristics of nearby cities. In Table A7, we create an instrument that does not take into account plants placed either in a 200km or a 400km radius of a given city, and find that, if anything, our results are more precisely estimated.

We may also be concerned that newly built plants are subject to more pollution regulation as the Chinese government paid more attention to environmental issues in recent years. So, in Table A8 we only restrict our sample to old plants, and find similar effects. In Table A9, we drop large coal producing provinces to address the concern that coal-fired plants may locate near coal production. In Table A13 we show that baseline population, GDP and electricity consumption do not predict future upwind plants, or future iterations of our IV. The results suggest that it is not that policy makers avoid richer, influential cities when building plants, and that plants are not built in areas that have a higher need for electricity at baseline, perhaps as most electricity is directly supplied to the larger grid.

In Table A14, we extend this analysis by creating various 'placebo' instruments, artificially changing the wind direction and showing that these falsified instruments do not predict pollution levels nor migration decisions. 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 the variation in pollution driven by China's Huai river heating policy. As Chen et al. (2013) show, the heating policy 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.1.3 we examine the consequences of this policy. While we fail to find differential out-migration in response to pollution, we find stark differences in in-migration between skilled and unskilled workers in response to air quality differences.

We summarize our analysis of the different sources of variation in Section 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.2, we turn our attention to studying different model specifications, subsamples, and checking the robustness of our estimates to different controls. First, in Appendix A.2.1, we use an individual-level longitudinal panel data and a different definition of migration to replicate our results. The data are constructed using the China Labor-force Dynamic Survey. The longitudinal panel allows us to track individuals' spatial sorting over time, and control for city and individual-level characteristics. Importantly, we define migration to be an indicator for whether or not an individual changed their city location between years, regardless of whether they change their *hukou* location or not. As reported in Table A16 and A17, we again find a similar empirical pattern–high-skilled workers are more likely to leave polluted cities relative to their low-skilled counterparts.

We then study the effects of cumulative pollution exposure in Appendix A.2.2. The results shown in Table A19 and A20 indicate that pollution exposure spread over a longer time period has a larger impact than shorter time frames. Again, workers with high education attainment are more sensitive to cumulative pollution than those without.

In Appendix A.2.3, we examine different samples and perform heterogeneity analysis. In Table A21, we disaggregate education levels into more categories and see a sharp education gradient in out-migration responses: those with more education are more responsive. 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

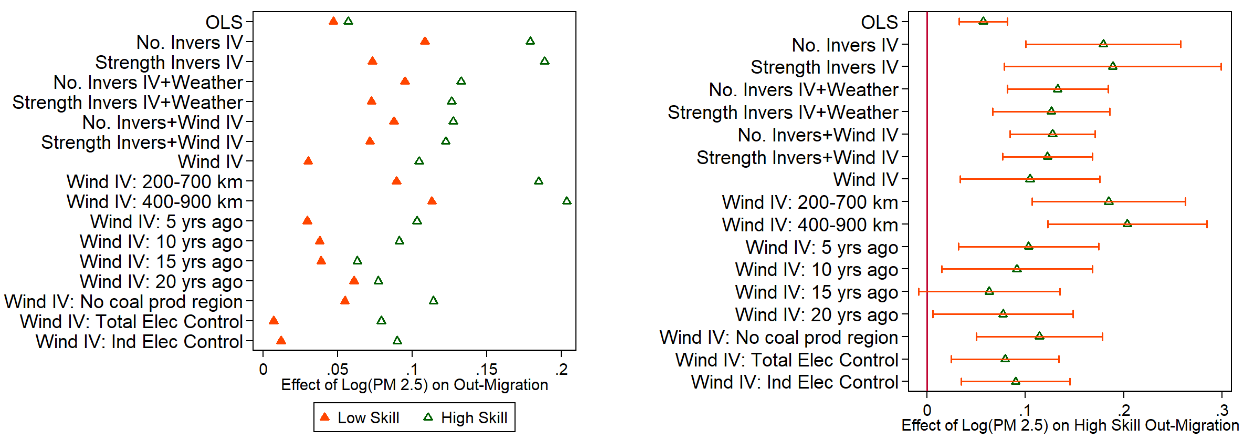
A23, we find that the youth are more responsive to pollution when making location choices, while Table A24 studies heterogeneity across rural-urban status.

We next employ an alternative measure of local air quality. Table A25 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.

Finally, Appendix A.2.4 highlights robustness to a long list of controls, including skill-distribution indicators (first two columns in Table A26), baseline economic indicators (third and fourth columns in Table A26), as well as industrial pollutants emission (last two columns in Table A26). 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



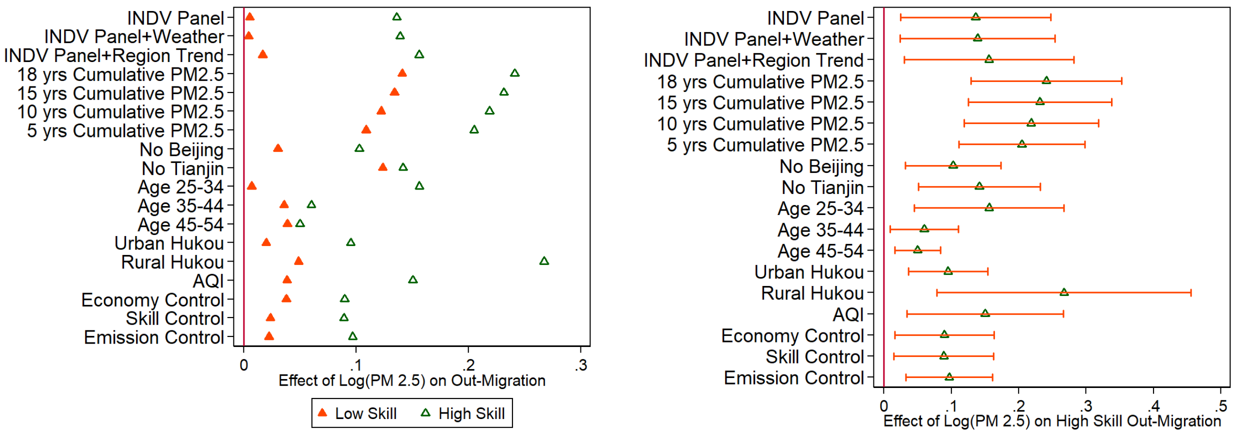
Notes: Summary of results using different sources of variation. We compile coefficients from different specifications. On the left we show both the coefficients on high and low skilled workers. On the right, we concentrate on high skilled workers, and include 90% confidence intervals.

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 the effects of Log PM2.5 across alternative sources of variation. We first show OLS results before going into various instrumental variable estimates. Our empirical pattern are robust to instrumenting for air pollution using the number ("No. Invers" for short, in the figures) and the strength of thermal inversions ("Strength Invers", in the figures) as well as different versions of the wind direction and coal-fired power plants IV ("Wind IV", in the figures). We further combine wind and inversions instruments, the results are qualitatively similar.

The results estimated using thermal inversions IV hardly change when we account for local weather conditions. The IV estimation that rely on wind direction and coal-fired power plants was subject to a wide range of tests discussed, and so has many point estimates under display. For example, “Wind: 20 yrs ago” relies on plants built before 1995, 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 direction, distance to cities and access to coal. The “Wind IV 400-900 km” IV excludes any plants built within a 400km radius from a city and instead only captures plants built 400-900 km away. This is done to allay concerns related to the endogenous placement of plants in close proximity to cities.

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

Figure 9: Different Models, Samples and Controls



Notes: Summary of results using different models, samples and controls. 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 high-skilled workers, and include 90% confidence intervals.

In Figure 9 we summarize our results for alternative model specifications, samples and controls with the help of some figures. First, we employ an individual longitudinal panel and account for both individual and city fixed effects (“INDV panel”, in the figures). We still find a similar empirical pattern. Such patterns hardly change when further controlling for local weather conditions (“INDV panel+weather”, in the figures) and region-specific trends (“INDV panel+region trend”, in the figures). We next turn to the implications of accumulated PM2.5 exposure on migration decisions. The longer the period of pollution exposure, the stronger the response. We also use different samples to replicate our results and conduct heterogeneity analysis. Finally, we test the sensitivity of our results by adding a range of covariates. Consistently, the left-side graph shows how the effects on high-skilled worker migration is always larger than that for low-skilled out-migration. The right-side figure shows that the effect on high-skilled worker out-migration is strong and statistically different from zero.

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 cities. Skilled workers are going to be scarce in cities that they leave, raising their value in the labor market in such cities. Additionally, given the complementarity between skilled and unskilled workers, cities that lose skilled workers will have less productive unskilled workers. As such, cities that lose skilled workers would have higher skilled wages, lower unskilled wages, and therefore higher returns to skill.

Table 2: Pollution and Returns to Skill

	Dep variable: City-specific returns to a college degree			
	OLS		IV	
Log(PM2.5)	0.248**	0.453***	0.789**	1.570***
	(0.125)	(0.135)	(0.309)	(0.539)
Observations	130	130	130	130
City Controls	N	Y	N	Y
Weight	population 2000	population 2000	population 2000	population 2000

Notes: Standard errors at the 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 Shanghait, Tianjin and Shenzhen seaports.

Table 2 displays a simple empirical fact: wage returns are higher in cities that have more pollution. We estimate the city-specific Mincerian returns to skill using CLDS, and explore the relationship between air pollution and returns to skill. Panel A in Table 2 shows the OLS results. The impacts of pollution on returns to skill are positive and statistically different from zero. As presented in Panel B, the coefficient estimates of Log PM2.5 increase in magnitude when we employ the wind direction and coal-fired plants IV to address the endogeneity concern. 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.

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 makes it easier for skilled workers to move to certain cities and have access to jobs and public services. In Table C2 we show a few examples of such restrictions, highlighting how education levels can help one gain enough *hukou* points to be eligible. These 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. Second, 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 responsive to city level amenities and as such, respond to changes in amenity values, such as pollution levels. In Table C3 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 high school education. We find that those with more education are more likely to claim that the environmental issue in China is "terrible." In Table C4, we use the same survey to explore not just 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 some college or above education are meaningfully more concerned about environmental issues than those with less than high school education, and also than those with high school 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.[17] An additional source of geographic-specific returns to skill is driven by the fact that some cities have more skill-biased capital. Moving a skilled worker from a low wage (productivity) city to a high wage city will increase aggregate productivity. Relaxing *hukou* restrictions may improve welfare as individuals may no longer be stuck in low-wage,

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

high-polluting cities.

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 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 health 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 may vary across cities, and may depend on the lack of pollution Z_d, and agglomeration forces.[18] Capital is perfectly elastically supplied across cities at rental rate R^*.[19] Effective labor supply L_d depends on labor L_{sd} at each skill level s.

$$Y_d = A_d L_d^{\varrho} K_d^{(1-\varrho)} \quad where \quad L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E - 1}} \tag{4}$$

$0 < \varrho < 1$ 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.

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 $\theta'_{sd}(k_{sd}) > 0$.[20] Notice, it could also capture the fact that some cities have policies that raise wages for skilled 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:[21]

$$\log w_{sd} = \log \left(\frac{\partial Y_d}{\partial L_{sd}} \right) = \frac{1}{\varrho} \log A_d + \log \widetilde{\varrho} + \log \theta_{sd} + \frac{1}{\sigma_E} \log L_d - \frac{1}{\sigma_E} \log L_{sd} \,, \tag{5}$$

[18] Keeping with convention in the literature on amenities, higher Z_d means more amenities, so less pollution.

[19] The perfectly elastic capital assumption is not essential and can be relaxed. See Appendix E.2.

[20] For completeness, one can explicitly model skill-biased capital within the nested CES framework and show how incorporating it do not affect the qualitative predictions. See Appendix E.2.

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

where $\log \widetilde{\varrho} \equiv \left(\frac{1-\varrho}{\varrho}\right) \log\left(\frac{1-\varrho}{R^*}\right)$ is common across all cities and workers.[22] 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.

First, A_d is the amount of TFP at the city level, which may raise average earnings at the city level. Second, θ_{sd} is the higher skill-biased 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. Third, earnings differ due to differences in the supply of more educated workers L_{sd}. As with any downward sloping demand curve, the more skilled workers there are, the lower the skilled wage. Yet, L_d also accounts for the complementarity effect, whereby an increase in the number of unskilled workers may actually raise the skilled wage.

Equation 5 is the (inverse) 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 with relatively more skilled workers, the skilled wage will be relatively lower. Whereas for regions with more skill-biased capital, the skilled wage 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 things. First, aggregate population, which in turn captures the size of the economy, industrial production and congestion. Second, the skill mix, which captures types of production (industry vs services), changes to local laws, and amenities produced as the skill mix changes.

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

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, from origin city o is given by:

$$V_{jsod} = \epsilon_{jsd} w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d exp^{-M_{sod}}\ , \tag{7}$$

where ϵ_{jsd} is a random variable measuring preferences for a specific city d by individual j. A larger ϵ_{jsd} means the worker is particularly attached to a city d. M_{sod} captures the migration costs between o and d, including *hukou* costs of migrating to the city for migrants. As such, migration costs vary by education level, and $M_{soo} = 0$ for natives. hp_d are housing prices, and

[22] For tractability, output is the numeraire. Housing is not traded across cities, and will have price effects across cities and skill groups. Since output is not skilled biased, a price of output would not change the skill-specific returns.

ν_s are 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_{sd} represents other non-pollution related skill-specific amenities. The compensating differential elasticity will be captured by γ_s, and varies by skill level.[23] Here, marginal workers are those that are indifferent across cities, and so 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_{jsd} are independently distributed and drawn from a multivariate extreme value distribution (Eaton and Kortum, 2002). The joint distribution of ϵ_{jsd} is given by:

$$F(\epsilon_{s1}, ...\epsilon_{sD}) = exp\left(-\sum_{d}^{D} \epsilon_{sd}^{-\eta_s}\right) , \tag{8}$$

where $\frac{1}{\eta_s}$ 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 migration costs (including *hukou* costs) are captured by ϵ_{jsd} and M_{sod} respectively. A person chooses city d over d' if $V_{jsod} > V_{jsod'}$. In Appendix E.1 we derive an expression for labor supply and average city-utility. The share of workers with skill s from city o that move to d is given by:

$$\pi_{sod} = \frac{\left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_{sd} exp^{-M_{sod}}\right]^{\eta_s}}{\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{sd'} exp^{-M_{sod'}}\right)^{\eta_s}} \tag{9}$$

The supply of workers of skill level s in city d additionally depends on the *hukou* population of city o, P_{os}.

$$L_{sd} = \sum_{o} P_{os} \pi_{sod} \tag{10}$$

Taking logs of Equation 9, we derive the labor supply curve:

$$log\ \pi_{sod} = \eta_s log\overline{V_{so}} + \eta_s \left(log\ w_{sd} - \nu_s log\ hp_d\right) + \eta_s log\ a_{sd} + \eta_s \gamma_s log Z_d - \eta_s M_{sod}\ , \tag{11}$$

where $\overline{V_{so}}$ captures the average (option-value) utility of being from city o.[24] Note that because of migration costs, utilities are not equalized across cities, and as such the term has an o 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

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

[24]As we show in Appendix E.1, we can derive: $\overline{V_{so}} = \left(\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{sd'} exp^{-M_{sod'}}\right)^{\eta_s}\right)^{\frac{1}{\eta_s}}$.

value of moving to a potentially desirable location falls. Therefore, as we show in Appendix E.1 that $\overline{V_{so}}$ depends on city-specific *hukou* restrictions.

From Equation 11 we see that $\frac{1}{\eta_s}$ 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.

6.3 Productivity, Agglomeration and House Prices

We consider the role played by agglomeration economies, and allow for pollution to directly affect productivity. In Equation 12, $\bar{A}_d$ is 'exogenous' city-level productivity (fertile soil, rivers, land etc.). ϕ_1 determines the elasticity of pollution with TFP. If $\phi_1 < 0$, then pollution is associated with more productivity. If $\phi_1 > 0$, pollution lowers the productivity of all workers. We expect the share of skilled workers to raise TFP levels in the city via non-excludable innovation (Arrow, 1962), such that $\phi_2 \geq 0$.

$$A_d = \bar{A}_d Z_d^{\phi_1} L_{sd}^{\phi_2} \tag{12}$$

While we do not explicitly need to model housing supply, like in Moretti (2011), we assume a simple housing market of the form $hp_d = (L_{sd} + L_{ud})^{\psi_3} \frac{L_{sd}}{L_{ud}}^{\psi_4}$, where more people in the city raise house prices, and relatively wealthier residents raise them further.

6.4 Equilibrium and Elasticities

Equations 4- 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. To be specific, the model is characterized by a set of exogenous factors: city level producitivities $\bar{A}_d$, total populations of the skilled and unskilled $\bar{L_u}$ and $\bar{L_s}$, migration costs M_{sod}, amenities α_{sd}, skill-biased capital θ_{sd}, and exogenous components of pollution $\bar{Z}_d$; and a set of parameters $(\sigma_E, \gamma_s, \eta_s, \phi_1, \phi_2, \psi_1, \psi_2, \psi_3, \psi_4)$, that together determine the quantities $A_d, Y_d, L_{sd}, L_{ud}, Z_d$ and prices w_{sd}, w_{ud}, hp_d.

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 $k = \{s, u\}$, as in Equations 5, 9 and 10. Total population of the skilled and unskilled in the country is the sum of the city level populations (or, $\bar{L_s} = \sum_d L_{sd}$ and $\bar{L_u} = \sum_d L_{ud}$). The sum of shares of migrant and non-migrants sums up to one, or $\sum_d \pi_{sod} = \sum_d \pi_{sod} = 1$. Output 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 city $Y = \Sigma_d^D Y_d$.

When bilateral migration costs are present we make a few other standard assumptions that help meet sufficient conditions for the existence of a spatial equilibrium (Allen et al., 2020): M_{sod} are finite, the graph of the matrix of costs is strongly connected, and they are quasisymmetric.[25] Our model contains congestion forces (such as pollution and house prices) and agglomeration (effects on TFP). An equilibrium is unique if congestion forces are at least as large as the agglomeration forces.[26]

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 labor supply curve, such that the difference in wages is the compensating variation (in partial equilibrium). In (general) equilibrium though, skill-based wages 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_s (local ties / preferences, 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

These are two important model results that we have already confirmed empirically. First, that skilled labor falls as the amount of air pollution increases. Second, that the returns to skill rise as pollution increases. Yet, 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_s). Instead, in general equilibrium it is the result of future migration changes as wages change (given the size of η_s) 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.

We estimate the following city-level parameters: $\{\theta_{sd}, \alpha_{sd}, M_{sod}\}$ and aggregate elasticities: $\{\sigma_E, \eta_s, \gamma_s, \psi_1, \psi_2, \psi_3, \psi_4, \phi_1, \phi_2\}$ based on city-level relationships for a set of large cities for which we have consistent data on all the variables across years. In our preferred estimates we weight by population and control for the city characteristics as in our earlier results, but show robustness to alternative specifications. We find estimates of these parameters to be similar

[25] The connectivity assumption simply implies that there is a sequential path of finite bilateral migration costs that can link any two cities o and d. The quasi-symmetry assumption, which is not entirely necessary for existence (but does aid the solution), simply says that one portion of the costs is symmetric. That is, $M_{sod} = M_{so}M_{sd}\widetilde{M_{sod}}$, and $\widetilde{M_{sod}} = \widetilde{M_{sdo}}$. So in moving from Tianjin to Beijing, there may be a component of the cost that is Beijing specific (say, related to Beijing *hukou* policy), a component related to leaving Tianjin (say, its large airport), and a component that is Beijing-Tianjin specific (say, the distance between the two, or number of train connections). This last bilateral component is assumed to be symmetric for ease of proving the existence of an equilibrium.

[26] That is, the parameters ψ_1, ψ_2 and ψ_3 that determine congestion are meaningful in magnitude, relative to ϕ_1 and ϕ_2 that drive agglomeration.

what has already been estimated in the literature. While we may calibrate our model from parameters estimated by other work, we feel more confident in using parameters estimated within our own model and data. To causally estimate these parameters we derive additional sources of variation from other changes in policy.

7.1 Labor Demand Curve: Estimating σ_E

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 from Equation 5, where within a city d, the size of a city's output (and other city-level characteristics) are differenced out, as in Equation 13:

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

The parameter σ_E, can be estimated from this relative labor demand curve, as exogenous shifts in relative labor supply $log\frac{L_{sd}}{L_{ud}}$ can help trace out the relative labor demand curve and identify the slope, $1/\sigma_E$.

As the relationship between the number of workers and wages is determined in equilibrium, we use a two-staged least squares (2SLS) relationship, where, as we have established, pollution is a shifter of labor supply. As pollution shifts the labor supply curve it traces out the labor demand curve. We estimate Equation 13, with the following first stage:

$$log\frac{L_{sd}}{L_{ud}} = \alpha_0 + \alpha_1\ log\ PM2.5_d + \varepsilon_{1d} \tag{14}$$

Table 3: Estimating Labor Demand Elasticities

IV:	Coal+Wind IV		No. of inversions		Inversion strength	
	Ln $\frac{w_s}{w_u}$	Ln $\frac{L_s}{L_u}$	Ln $\frac{w_s}{w_u}$	Ln $\frac{L_s}{L_u}$	Ln $\frac{w_s}{w_u}$	Ln $\frac{L_s}{L_u}$
Log(PM25)	1.000**	-1.238**	0.703***	-1.458***	0.850***	-1.681***
	(0.438)	(0.617)	(0.244)	(0.366)	(0.316)	(0.554)
Observations	130	130	130	130	130	130
City Controls	Y	Y	Y	Y	Y	Y
Weather	N	N	Y	Y	Y	Y
σ_E	1.24		2.07		1.98	

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 at the city level are reported in parentheses. All regressions weighted by the population in 2000.

We perform such an exercise in Table 3. We estimate Equation 14 in the columns where our outcome of interest is the relative stock of workers Ln $\frac{L_s}{L_u}$, capturing the net migration for all types of workers (whether or not they changed *hukou* location). Together with the help of the columns, where the outcome is $\text{Ln}\frac{w_s}{w_u}$, we can estimate the relationship in Equation 13. We take the ratio of the IV relationship for quantities of workers, and the wages of workers. For instance, in the last two columns, we find that $\sigma_E = 1.238/1 = 1.238$. This suggests that the elasticity of substitution across skill levels is 1.238, an estimate that is also close to the estimates found in the US (Card and Lemieux, 2001). As such, using calibrated elasticities-of-substitution from the literature produce similar model counterfactual exercises below.

While σ_E is identified by the change in relative wages in response to changes in relative skill-shares, γ_s will be identified below by the initial shift in workers in response to pollution. Even though γ_s 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 (via σ_E), which in turn affects future migration (η_s), 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.

7.2 The Labor Supply Curve: Estimating (η_s, γ_s) and $\{M_{sod}\}$

The labor supply curve in Equation 11 captures bilateral migration flows between pairs of cities, as a function of real wages at destinations, pollution levels at the destination, and migration costs between origins and destination. First, we assume that migration costs have the following functional form:

$$M_{sod} = \lambda_{1s} log\ Dist_{od} + \lambda_{2s} \left(\mathbb{1}_{Migrant_{od}} \times hukou_{sd}\right) \ , \tag{15}$$

where $log\ Dist_{od}$ is the log of the distance between cities o and d, $\mathbb{1}_{Migrant_{od}}$ is an indicator for whether the $o \neq d$, and $hukou_{sd}$ is the *hukou* index, derived from Zhang et al. (2018). We use the inverse hyperbolic sine of migration distance to capture the physical and physiological costs associated with being far from ones origin city. We use the interaction of hukou index and migration status indicator to capture institutional migration costs due to hukou restrictions. 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). Specifically, 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 is a more restrictive policy. The index is highest for Beijing, followed by many of the other major cities.

Substituting Equation 15 in 11, we derive our estimation equation for labor supply:

$$\begin{aligned} log\ \pi_{sod} =& \eta_s log\overline{V_{so}} + \eta_s \left(log\ w_{sd} - \nu_s log\ hp_d\right) + \eta_s\gamma_s log Z_d + \mathbf{X}\beta_x \\ & - \eta_s\lambda_{1s} log\ Dist_{od} - \eta_s\lambda_{2s}\left(\mathbb{1}_{Migrant_{od}} \times hukou_{sd}\right) + \epsilon_{3sod} \ , \end{aligned} \tag{16}$$

where, $\varepsilon_{3sod} = (\eta_s log\, a_{sd} + \varepsilon_{2sod})$, the residual includes differences in destination city amenities, and other idiosyncratic features determining bilateral flows. Below, we describe how we derive amenities from the residual by inverting the model. In our estimation, we include origin-city-by-skill fixed effects to control for $\eta_s log\overline{V_{so}}$. We include controls, $\mathbf{X}$, including a city's $hukou_{sd}$ index. As such, the interaction with migration status, allows us to isolate the part of *hukou* index that affects migration costs, while controlling for the index accounts for any differences in city level characteristics (correlated with the *hukou* index), that affect migrants and non-migrants in a similar fashion. Like before, we instrument for pollution using thermal inversions.

We estimate η_s using instruments that shift out the demand curve by skill, tracing out the supply curve. Labor demand shocks, on the other hand, change θ_{sd}, and shift out the 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 instrument for changes in real wages by skill.

We derive variation 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. Since we have two labor supply elasticities η_s, for the skilled and the unskilled, we define two instruments: one for the real skilled wages, and another for the real unskilled wage. Both instruments are widely used by researchers to study the impact of Chinese exports on US manufacturing employment, namely by Pierce and Schott (2016) and Autor et al. (2013).[27]

The first instrument, the NTR gap, popularized by Pierce and Schott (2016) relies in the changes to the Normal Trade Relations (NTR) tariffs. Prior to joining the WTO, the US Congress needed to continually renew the preferential NTR tariffs bestowed upon China. Joining the WTO reduced the renewal uncertainty (captured by the NTR Gap) defined to be the difference between the non NTR tariff and the NTR tariff. While other work examines the impacts on US employment using US data, we use the same instrument to study what happens to internal migration in China as real wages change across cities.

We create city-level uncertainty measured by looking at the weighted sum of industry i's export shares EX_{di} in 1997, interacted with the industry-level NTR gaps:

$$NTR\ IV_{ud} = \sum_i \frac{EX_{di}^{1997}}{\sum_j EX_{dj}^{1997}} \times (nonNTR\ tariff_i - NTR\ tariff_i) \tag{17}$$

Comprehensive details about this trade shocks can be found in Khanna et al. (2020), where we show that the NTR gap instrument is better at predicting changes to real unskilled wages, rather than skilled wages, possibly as the industries that benefited most from such tariff changes

[27] We perform robustness checks surrounding these instruments in Khanna et al. (2020). Recent developments in the shift-share literature discuss additional tests, such as tests for pre-trends, baseline share correlations, and standard error corrections. We perform these tests, noting that we rely on the assumption that in our case, the 'shifters' are exogenous (as in Borusyak et al. (2018) and Adao et al. (2019)), rather than the 'shares' being exogenous (Goldsmith-Pinkham et al., 2020).

were more likely to hire unskilled labor. We use this as an instrument for $log\ w_{ud} - \nu_s log\ hp_d$. We 'deflate' all our wages by local house prices, using yearly average data on housing rents from the Xitai Real Estate Big Data depository.[28]

For skilled real wages we leverage variation from the World Import Demand (WID) for skilled industries. Following Autor et al. (2013), we use world-import demand shocks by industry, and weight those by initial export shares to create a city exposure measure. Since we aim to create an instrument for skilled-wage, we use the share skill-intensive industries.[29]

$$WID_{sd} = \sum_i \frac{EX_{di}^{1997}}{\sum_j EX_{dj}^{1997}} \times \left(\frac{World\ IM_{i,2015} - World\ IM_{i,2004}}{World\ IM_{i,2004}} \right) \tag{18}$$

Table 4: Estimating Labor Supply Elasticities

IV-2SLS Labor Supply	Low Skill Workers $Log\ \pi_{uod}$		High Skill Workers $Log\ \pi_{sod}$	
$Log(PM2.5)_d$	-0.0427***	-0.0488***	-0.506***	-0.513***
	(0.0090)	(0.0114)	(0.0958)	(0.0829)
$Log(Real\ Wage)_d$	1.012***	1.126***	1.301***	1.024***
	(0.269)	(0.318)	(0.251)	(0.170)
$Log(Distance)_{od}$	-0.0754***	-0.0783***	-0.0308***	-0.0404***
	(0.0112)	(0.0129)	(0.0052)	(0.0046)
$Hukou\ Index_{sd} \times Migrant_{od}$	-0.923*	-0.852*	-3.489***	-3.061***
	(0.487)	(0.501)	(0.876)	(0.758)
Observations	13,570	13,570	13,570	13,570
Pollution IV	No. thermal	Strength thermal	No. thermal	Strength thermal
Wage IV	NTR IV	NTR IV	WID IV	WID IV
Controls	Yes	Yes	Yes	Yes
Hukou City FE	Yes	Yes	Yes	Yes
First stage F-stat	19.45	15.16	30.97	70.07

The NTR IV is the weighted average of the NTR gap, where the weights are the baseline industry level export shares (Pierce and Schott, 2016). The NTR gap is measured as the gap in Normal Trade Relation (NTR) tariffs and the non-NTR tariffs. The WID IV is the weighted average of the world import demand, where the weights are the baseline skill-intensive share of industries Autor et al. (2013). The measure of $Hukou\ Index_d$ varies across cities and skill level (Zhang et al., 2018). We control for this (non-interacted) measure of the hukou index. All regressions also control for temperature, humidity, sunshine duration, and wind speed, as before when using thermal inversions as an IV. the High skill workers are those whose highest degree is some college or above, and low skill workers are those whose highest degree is high school or below. Standard errors clustered at the destination city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

[28]Further details and data sources at http://www.cityre.cn/credata.html. The data has been collected since 2005, and covered 337 cities in China, in collaboration with the China National Bureau of Statistics, and the China National Development and Reform Commission. We compare these data to the purchase price of residential properties from the China statistical yearbook, and find a correlation of 0.93.

[29]We label industries as skill intensive if they are above the median in the ISIC industry data. The skill share is the share of skilled workers in the industry, based on the Annual Survey of Industrial Production (ASIP) available in 2004. We aggregate the firm data into 4-digit ISIC industries. For instance, in ISIC 1810, 5% of the labor force is "skilled". We construct our measure using the Indonesian manufacturing census (Amiti and Freund, 2010), so as to ensure no confounding effects of using the same sample to construct our skill-intensity measure and regression estimation.

Table 4 allows us to estimate η_s and γ_s for each skill group. Across the columns we vary the skill groups and the instruments used. The skill-biased trade shocks will raise the demand for some occupations more than others. This changes the wages by city and skill group in response to the trade shocks, and helps us identify the labor supply curve in response to changes in wages. We estimate η_s from Equation 16 leveraging variation that arises from China's accession to the WTO. Our estimates suggest that $\eta_u = 1.012$ and $\eta_s = 1.301$. These are similar to estimates of labor supply elasticities from Tombe and Zhu (2019).

The coefficients on $Ln(PM2.5)_d$ provide us with estimates of $\eta_s \gamma_s$. The γ_s parameters capture the marginal utility of clean air, and vary by skill level. Along with our estimates of η_s, we know that $\gamma_s = 0.38$, and $\gamma_u = 0.042$, once again reiterating how the skilled are more sensitive to air quality than the unskilled. When comparing the labor supply elasticities with respect to wages and to pollution levels, we can see that both types of workers are far more responsive to changes in wages than they are to pollution levels.

Workers are also respond to migration costs. Since hukou regulations may be correlated with city-level unobservables and industrial policy, we control for the direct (un-interacted with migration indicators) measure of the *hukou* index. As such, we control for city level characteristics that may be correlated with the attractiveness of cities, and are relying on only the interaction with migration status for identifying the response to *hukou* restrictions.

Table 4 shows that migration between origin o and destination d is less likely to occur over longer distances, and if there are more *hukou* restrictions.[30] While skilled workers are less sensitive to distances, they are more sensitive to *hukou* restrictions. Despite the fact that skilled workers are more responsive to *hukou* restrictions, they face much lower level of restrictions. The responsiveness perhaps reflects higher preferences for access to amenities (like housing permissions and children's schooling) that are only obtainable via accessing local *hukou*.

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). As such, 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 Measuring Amenities and Productivities $\{\theta_{sd}, \alpha_{sd}, \bar{A}_d\}$

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:

$$\frac{w_{sd}L_{sd}}{w_{sd}L_{sd} + w_{ud}L_{ud}} = \frac{\theta_{sd}L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}}}{\theta_{sd}L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} + (1-\theta_{sd})L_{ud}^{\frac{\sigma_E - 1}{\sigma_E}}} \tag{19}$$

We plot the city-level distribution of θ_{sd} in Figure 10. Beijing and Shanghai have the highest

[30]We use the inverse hyperbolic sine of distances so as to include 0 distances.

amount of skill-biased capital in the country, which almost double the amount of the median city, many of which lie in less urbanized areas.

Figure 10: Distribution of θ_{sd} across cities, from Equation 19

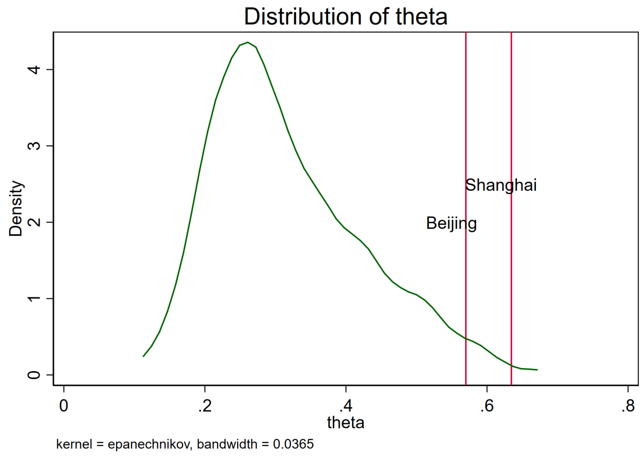


Table 5: Pollution, Population and TFP

	Agglomeration	Congestion Forces	
	Log(TFP)	Log(PM 2.5)	Log(House Price)
Log(PM 2.5)$_d$	-0.0512		
	(0.271)		
Log L$_{sd}$	0.0947*		
	(0.0542)		
Log Population$_d$		-0.00873	0.258**
		(0.138)	(0.126)
Log (L$_{sd}$/L$_{ud}$)		0.134*	0.415***
		(0.0809)	(0.0547)
Observations	120	119	119
Weather Amenities	Yes	Yes	Yes
First stage F-stat	16.23	12.36	12.36

Notes: The first column estimates the relationship between PM 2.5, number of skilled workers and TFP. We use thermal inversions as an instrument for PM 2.5, and leverage the higher education expansion instrument to identify the effect of the number of skilled workers. For our congestion forces we use instruments for population and the skill ratio that we describe in the text. All our regressions use the 2000 population as weights. We control for weather amenities (temperature, humidity, sunshine duration and wind speed), as in our other specifications. For TFP regressions we control for distance to seaports and region fixed effects.

We follow the literature (Ahlfeldt et al., 2015; Bryan and Morten, 2018) and derive non-pollution amenities as a residual from the labor supply curve, Equation 16. Intuitively, bilateral migration flows between cities must be driven by wages, house prices, pollution, migration costs, and other residential amenities. As we measure, and account for the effect of all other components, the remaining portion of the flows are due amenities. We take the residual for each skill-specific regression estimated by Equation 16, and measure amenities using destination fixed effects $\widehat{\epsilon_{3sod}} = \alpha_{sd} + \epsilon_{2sod}$, where $log\ a_{sd} = \widehat{\alpha_{sd}}$.

To quantify changes to output, we first measure of TFP $\bar{A}_d$, which captures features of the local area (e.g, land quality). We follow the literature (Ahlfeldt et al., 2015) and measure TFP as the city-level aggregate residual from output. After accounting for the optimal (unbiased) capital flows, output is simply: $Y_d = A_d^{\frac{1}{\varrho}} \left(\frac{1-\varrho}{R^*}\right)^{\frac{1-\varrho}{\varrho}} L_d$. Using our estimates for σ_E and θ_{sd}, we are able to create a measure of L_d. We then invert the model to derive A_d.[31]

7.4 Estimating Agglomeration and Congestion Forces (ϕ and ψ)

We use Equation 20 to study how our measure of TFP correlates with pollution as instrumented with the thermal inversions instrument for pollution.

$$\widehat{\frac{1}{\varrho} \log A_d} = \log \bar{A}_d + \phi_1 \log Z_d + \phi_2 \log L_{sd} + \varepsilon_{4d}\ , \tag{20}$$

where ϕ_1 is the elasticity of pollution with aggregate TFP.[32] Notice, A_d may capture other drivers of city-level TFP, like land, housing supply, innovation. Equation 20 also allows the number of skilled workers 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.[33]

In Table 5 we estimate ϕ_1 using Equation 20, and leveraging our instrumental variables strategy for PM 2.5 emissions. While not precisely estimated, we conclude that $\phi_1 = -0.051$. (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. As such, our estimates lie within estimates from other work.

In addition, we estimate estimate the impact of a changing skill share on TFP to capture in agglomeration effects. We derive this variation from a university expansion in China, where a large number of new universities were built at the turn of the century in certain cities. These cities saw a sharper growth in the number of college graduates.[34] As such, we instrument for

[31] $\log \bar{\varrho} \equiv \left[\left(\frac{1-\varrho}{\varrho}\right) \log\left(\frac{1-\varrho}{R^*}\right)\right]$ is common across all cities and workers. ϱ is nothing but the labor share of income in the city.

[32] TFP may also be positively associated with pollution. This may generally be true if more production leads to more pollution, but we are using an IV for pollution here.

[33] 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.

[34] 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.

the number of skilled workers skill using the higher education expansion in China over the early 2000s. Figure C2 describes this event and shows a lack of pre-trends in cities that received the college expansion.

We aggregate the number of college graduates by city for the first graduating cohort of the expansion (2001-2005). Our specific instrument is the difference in the number of college graduates in 2005 and in 2001. Our estimate of ϕ_2 suggest a meaningful agglomeration effect.

Our remaining parameters include how changes in the number of workers, and skill share affect the amount of pollution and house prices in each city. In order to estimate ψ_1 we require variation in the skill-ratio that is not driven by the air quality. As such, we leverage the college expansion policy (dividing it by the baseline population in 2000 to get a ratio) to estimate the effect of a change in skill ratios.

Simultaneously, we estimate how changes in total population affect air quality, captured by parameter ψ_2. We leverage push factors from migrant-sending cities to estimate the effect of changing populations. We weight the growth in out-migration from all provinces in China, with the baseline shares of migrants from each province that came to the city in 1990. When calculating the out-migration growth (between 2005 and 2015), we exclude flows to the destination city itself so as to not capture labor demand changes at the destination. We create the population instrument at the city level:

$$Population\ IV_{d,2015} = \sum_p \left(\frac{Migrants_{dp,1990}}{\sum_d Migrants_{dp,1990}} \times \left(\sum_{d' \neq d} \Delta Migrants_{d'p,2015-2005} \right) \right) \tag{21}$$

Note, that this is different from the instruments introduced by (Card, 2001) and described by Jaeger et al. (2018). As we have rich data on outflows of migrants from provinces, we can leverage 'push factors' from sending regions, rather than rely on data solely on inflows into destinations as previous work does. As such, we do not need to assume that the baseline shares are exogenous, but rather in the shift-share framework described by Borusyak et al. (2018) identification entails that the forces driving the outflows from provinces to all other cities are not associated with city d specifically. Yet, this instrument does allow us to use the strengths of the (Card, 2001) framework where baseline migration determines networks that may in the future determine migration flows when there is more out-migration from origin provinces.[35]

Table 5 describes the effect of population and skill ration on pollution. We find that our instruments are strong, as evidenced by the first stage. Our two-staged least squares estimated tell us that ψ_1 is approximately zero, and $\psi_2 = 0.134$. Increases in population raise the amount of pollution in a city, and relatively more skilled workers raise pollution levels even more.

Finally, in an extension of our exercise, we consider how house price changes may affect

[35]Indeed, before 2000 internal migration was highly regulated, and driven by policies like the 'sent down youth' (Kinnan et al., 2018). Since 2000, migration costs have fallen substantially (Tombe and Zhu, 2019), allowing us to estimate the impacts of subsequent growth.

Table 6: Summary of Parameter Estimation

Parameter	Value	Definition	Identifying Variation
σ_E	1.24	Skill-elasticity of substitution Relative labor demand elasticity	Pollution-driven geographic sorting Changes in skill ratio affect wages
η_s	1.301	High-skill labor supply elasticity	World Import Demand trade shocks
η_u	1.012	Low-skill labor supply elasticity	NTR Gap trade shocks post WTO
γ_s	0.506	High-skill migration response to pollution	Pollution-driven response by skill
γ_u	0.0427	Low-skill migration response to pollution	Pollution-driven response by skill
θ_{sd}	[0.15,0.64]	Skill-biased capital	Skill-labor share in wage bill
ψ_1	0.134	Pollution response to changing skill ratio	University expansion
ψ_2	-0.0	Pollution response to population	Out-migration from origin provinces
ψ_3	0.248	House price response to skill ratio	University expansion
ψ_4	0.415	House price response to population	Out-migration from origin provinces
ϕ_1	-0.0512	TFP response to pollution	Pollution IV affect TFP residual
ϕ_2	0.097	TFP response to skilled workers	University expansion
λ_s	3.489	High skill response to hukou index	Skill-biased Hukou index
λ_u	0.923	Low skill response to hukou index	Skill-biased Hukou index

Notes: We summarize the parameter estimation using different instruments in this table. The values are from Tables 3-5 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.

our estimates. To do so, we need to estimate ψ_3, the elasticity of house prices with respect to population, and ψ_4, with respect to the skill ratio. We leverage the same instruments and find, $\psi_3 = 0.258$ and $\psi_4 = 0.415$. As such, a larger population raises house prices, and these prices rise substantially more when there are relatively more skilled workers in the city.

In Table 6 we summarize all these parameter values and sources of variation. Given our estimated parameters that determine GDP, Y_d, we can create a model-predicted measure of GDP based on our estimates.

7.5 Model Solution and Validation Exercises

Output in city d depends on the set of parameters: $\{\theta_{sd}, \sigma_E, \eta, \phi_1, \phi_2, \psi_1, \psi_2, \psi_4\}$, a set of 'endogenous' quantities: $\{Y_d, A_d, L_{sd}, L_{ud}, Z_d\}$, and 'exogenous' quantities: $\{\bar{A}_d, \bar{Z}_d, M_{sod}, \bar{L_u}, \bar{L_s}\}$. Prices, $\{w_{sd}, w_{ud}, hp_d\}$ are determined in equilibrium, with the output being the numeraire. Changes in exogenous pollution levels $\bar{Z}_d$ will affect the location of workers and TFP, thereby changing output Y_d in the city, and in other cities as well.

We solve the model starting with the list of parameters and exogenous quantities, and a set of initial conditions for the endogenous variables. After estimating the parameters, we no longer use information from the endogenous variables to solve the model. The primary market clearing condition is the labor market equilibrium, since for tractability, we do not model the floor space market, or demand for output. We pick different starting points, beginning in the vicinity of the observed equilibrium, but vary it by as much as changing the starting value by 20% for each endogenous variable. For this range of starting values, the model converges to the same unique equilibrium.[36]

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, but are not used when solving for model equilibrium. 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.

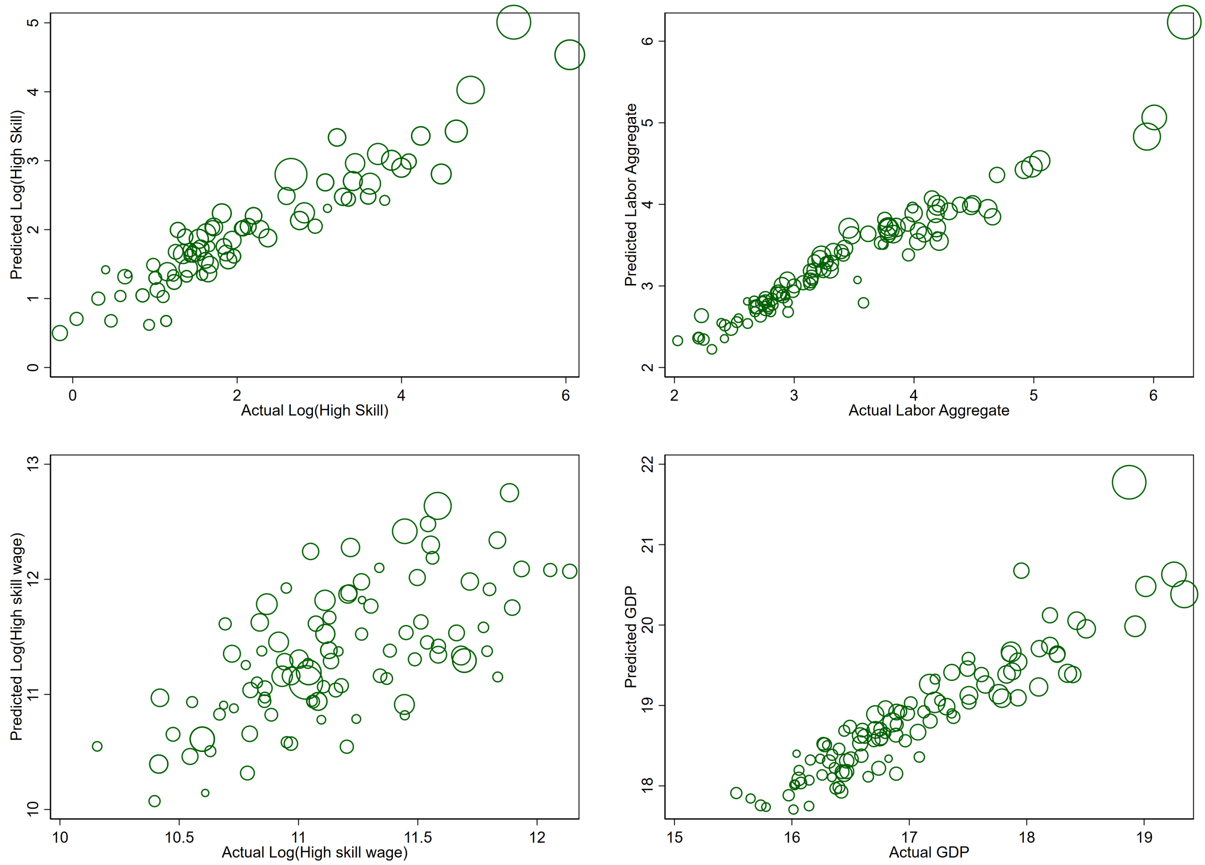
8 Counterfactuals: The Gains from Re-Allocation

There are three mechanisms through which pollution can result in productivity losses. First, and the one we highlight, skilled workers leave productive cities to avoid pollution. Reducing pollution in high productivity cities will induce skilled workers to relocate to cities with higher skill wages (and productivity), which in turn increases aggregate output. Importantly, given that costs of migration are higher for the unskilled, and as the unskilled are less responsive to poor air quality, the unskilled do not leave with the skilled. As we estimate the skilled and unskilled to be complements in production, this creates a mismatch on where they are located, reducing aggregate productivity.

Second, pollution further reduces the agglomeration of skilled workers in some productive cities, lowering productivity. Third, pollution has a direct effect on productivity through affecting workers' health. In this section, we conduct counterfactual exercises to quantify the role played by air pollution in determining the productivity losses via these three mechanisms. To understand the role played solely by the health channel, we prohibit workers from changing their location choice as a consequence of changing pollution levels. To shut down the agglomeration forces, we set the ϕ_2 parameter in the TFP equation to be zero, so as to not allow for

[36] This does not necessarily imply the equilibrium is globally unique. The existence of multiple equilibria often depends on the relative strength of the agglomeration and congestion forces. More skilled workers raise TFP (via ϕ_2), yet may lead to more congestion, captured by higher house prices (via ψ_3 and ψ_4) and more pollution (via ψ_1 and ψ_2), which in turn may lower TFP (via ϕ_1). Given the meaningful congestion forces we may expect a unique equilibrium.

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. Bubbles are weighted by populations in the year 2000.

any changes to city productivity.

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.

There are two types of pollution control policies we consider in our counterfactuals. First, we consider counterfactuals where we change the steady-state level of pollution Z_d in a city. This can be thought to be similar to policies where cities are set explicit pollution targets that they must meet regardless of how they do it (like, invest in greener technology). In the second type of policy, we relocate the exogenous component of pollution only, $\bar{Z}_d$. This policy is similar to relocating nearby coal-fired plants. However, relocating nearby plants may not effectively change equilibrium amounts of pollution in the city, as when more residents enter the city, and production patterns change, overall pollution levels may once again rise. This distinction between controlling steady-state levels of pollution and the exogenous component of pollution come not just with differing costs and policy prescriptions, but also differing consequences.

8.1 Changing Pollution in One City

Table 7 describes changes to pollution and migration policy in a highly industrialized and polluted city, Beijing. We first reduce the steady state PM 2.5 concentration by 50%, a policy akin to setting a pollution cap in Beijing. In the top panel we see that doing so raises GDP per worker by 11.11%. The health channel directly raises productivity by 3.5%. The pure relocation channel (without accounting for agglomeration economies) raises GDP per worker by a higher amount – 4.3%. If we allow for the fact that more high-skill workers imply more positive agglomeration spillovers, GDP per worker would rise by 7.35%. By comparison, the productivity gains through the indirect spatial sorting channel are substantially larger than the direct health benefits of air quality improvement.

Second, we examine what would happen if simply relocated the exogenous component of PM 2.5, allowing the steady state values of pollution to adjust when the population and skill composition of the city changes. This is a policy similar to relocating a nearby polluting powerplant to elsewhere in the country. In this case, we obtain very similar outcomes for GDP per worker and wages, perhaps reflecting that much of the pollution in Beijing was driven by the exogenous (say, location of power plants) component, $\bar{Z}_d$.

We then examine what would happen if we relax the *hukou* restrictions in Beijing to be the same as the median city in China, allowing workers to relocate, but not allowing pollution to change as a consequence. As reported in Table 7, the effect of lowering the *hukou* index in Beijing, depends on whether we change the skilled or unskilled *hukou*. When we lower the skilled *hukou* index, GDP per worker would rise by 13.37%, more than half of which is driven by the simple relocation channel, and the rest by the agglomeration forces. Since we do not allow pollution to change as a consequence of the population relocation, there are no health benefits. Lowering the unskilled *hukou* lowers GDP per capita, through a compositional change in the population – there are now more low-wage workers in Beijing.

If, in addition to reducing steady state PM 2.5, we also relax skilled hukou restrictions in Beijing to the extent that the stringency of hukou regulation is as same as its median level of all cities, then GDP would rise by 11.82% simply due to reallocation of workers, and by 20.76% given the agglomeration of skilled workers. These are much larger than the direct health benefits of clean air, 3.61%. Together the overall increase in GDP would be 25.13%. As a comparison, if we only relaxed hukou without reducing pollution, the rise in GDP would be smaller.

While changes in GDP per capita are informative of what happens to overall income levels

Table 7: The Productivity Effect of Reducing Pollution in One City

	Change in GDP per Worker (%)			
	Overall Change	Health Channel	Relocation	Relocation+Agglom
Reduce steady state PM2.5	11.513	3.613	4.396	7.625
Reduce exogenous part of PM2.5	11.114	3.504	4.229	7.352
Relax skilled *hukou*	13.373	0.000	8.214	13.373
Relax unskilled *hukou*	-7.102	0.000	-8.664	-7.102
Reduce PM2.5 & skilled *hukou*	25.125	3.613	11.817	20.762
Reduce PM2.5 & unskilled *hukou*	4.367	3.613	-3.908	0.728

Notes: In this counterfactual exercise we reduce the steady state amount of pollution in Beijing by 50% (row 1). We then reduce only the exogenous component of pollution by 50% (row 2). Next, we relax the *hukou* restrictions by skill level (rows 3 and 4) to the extent that the stringency of *hukou* regulation is as same as its median level of all cities. Finally (rows 5 and 6) we relax the hukou regulation to the same as its median level in China while reducing steady state pollution. Column 1 shows the gain to overall GDP per worker. Column 2 shows the component purely explained by the health-productivity channel. Column 3 through the pure relocation channel, and Column 4 also incorporates agglomeration forces as a consequence of relocation.

in Beijing, to understand the distributional consequences we look at the wages by skill group in Table 8. Once again, first we just reduce the steady state amount of pollution in row 1. When looking at wages by skill group, we see that overall skilled wages do not change. The improved productivity on the skilled by lowering pollution is counteracted by the reduction in wages as a consequence of an influx of skilled workers. Unskilled wages, on the other hand, rise sharply by 14.74%. Most of this is driven by the relocation channel, as when skilled workers enter Beijing, complementary unskilled workers become more productive. As a consequence average wages in the city rise by 11.11% (the same increase in GDP per worker). Reducing the exogenous part of pollution (row 2), once again produces similar impacts on wages.

Table 8: Distributional Consequences of Reducing Pollution in One City

	Skilled Wage (%)			Unskilled Wage (%)		
	Overall	Health	Relocate+Agglom	Overall	Health	Relocate+Agglom
Reduce steady state PM2.5	0.336	3.613	-3.162	15.232	3.613	11.214
Reduce exogenous part of PM2.5	0.338	3.504	-3.059	14.743	3.504	10.859
Relax skilled hukou	-7.110	0.000	-7.110	21.486	0.000	21.486
Relax unskilled hukou	14.466	0.000	14.466	-11.277	0.000	-11.277
Reduce PM2.5 & skilled hukou	-6.534	3.613	-9.793	40.156	3.613	35.269
Reduce PM2.5 & unskill hukou	14.607	3.613	10.611	1.964	3.613	-1.591

Notes: In this counterfactual exercise we reduce the steady state amount of pollution in Beijing by 50% (row 1). We then reduce only the exogenous component of pollution by 50% (row 2). Next, we relax the *hukou* restrictions by skill level (rows 3 and 4) to the extent that the stringency of *hukou* regulation is as same as its median level of all cities. Finally (rows 5 and 6) we relax the hukou regulation to the same as its median level in China while reducing steady state pollution. The first 3 columns show the effect on the wage of college educated workers, whereas the last 3 columns show the effects on the wage of the non college educated.

When we relax the *hukou* restrictions in rows 4 and 5 by skill group, we see that allowing in more workers of a particular skill group lower the wages of that skill group, while raising the productivity of the complementary skill group. For instance, relaxing the skilled *hukou* index lowers the skilled wage by 7.11%, but as skilled workers enter the city, unskilled workers become more productive and unskilled wages rise by 21.49%. When we combine relaxing the *hukou* regulation with changes to PM 2.5, we see that the overall changes are not simply the sum of the two counterfactuals conducted separately, highlighting the interplay between migration restrictions and pollution levels.

In Appendix Table D1 we examine the changes to GDP per worker, focusing solely on the main relocation effects, and shutting down any agglomeration or congestion effects (like, endogeonous changes to pollution, house prices and TFP), except the direct health-productivity benefits. The health and relocation channels similar to the main exercise in Table 7.

8.2 Relocating Pollution Away from Skill-biased Capital

We consider a second counterfactual where we keep the overall levels of pollution in the country to be the same, but simply relocate pollution from regions that have more skill-biased capital (high θ_{sd}) to regions with less skill-biased capital. This is again not as drastic a change as reducing pollution across the country, but perhaps a more feasible policy change. In this counterfactual, the level of PM 2.5 is directly mapped to the amount of skill-biased capital in the city. This does not necessarily change pollution in all cities, as a reasonable number of cities in the south-east have both clean air and high skill-biased capital already, whereas a fair number of low skill-biased capital industrial hubs already are heavy polluters.

In Table 9 we present the gains due to this relocation of pollution. In the first row, we relocate steady state levels of pollution. This is similar to a policy that sets pollution caps based on the amount of skill-biased capital in the city. Overall GDP in the country increases by 4.67%, whereby most of this increase is driven by the relocation of workers. The contribution of the health channel is only a 1.59% increase in GDP, while the relocation channel alone raises GDP by 2.07%. Agglomeration plays only a minor role.

When we relocate the exogenous part of pollution (shift a power plant away from a productive city, to a less productive one), the increase in GDP is smaller, reflecting that as people relocate to productive cities, pollution levels may increase in such cities and dampen the effects. As such, a policy that sets pollution caps may be more effective than a policy that relocates power plants. The contribution of the relocation and health channels are similar, and the agglomeration channel again matters less.

As a benchmark, in row 3, we relax the *hukou* restrictions in the top 35 (officially, first tier) cities to be the same as the median value of the *hukou* index in the country. This raises GDP by a similar magnitude to the pollution changes, but, by construction, is solely driven by worker relocation. When we lower overall migration costs in these top tier cities (row 4) to be the median cities value, the increase in GDP is a lot larger (about 6.94%). Lowering overall

Table 9: The Productivity Effect of Relocating Pollution Across Cities

	Change in GDP per Worker (%)			
	Overall changes	Health	Relocation	Relocate+Agglom
Relocate steady state PM2.5	4.671	1.589	2.069	2.097
Relocate exogenous part of PM2.5	3.873	1.452	1.422	1.403
Relax hukou	3.504	0.000	3.052	3.504
Relax overall mobility constraints	6.943	0.000	6.412	6.943
Relocate PM2.5 & relax hukou	7.736	1.589	4.656	5.076
Relocate PM2.5 & lower migration costs	11.599	1.589	8.065	8.656

Notes: In this counterfactual exercise we relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relocate only the exogenous component of pollution. In addition to such relocations of pollution, we also relax the *hukou* restriction in the 35 high tier cities to be the same level as the median city in the country (row 3). In row 4 we relax overall migration costs to the 35 high tier cities to be the same as the median city. Column 1 shows the overall gain to GDP. Column 2 shows the increase in GDP as a consequence of the health effects only. Column 3 shows the gain due to the re-allocation of labor channel only. Column 4 shows the gain to GDP after also accounting for changes in TFP due to changes in the agglomeration of skilled workers.

migration costs can be thought of as a policy mix of relaxing *hukou* restrictions and building more transportation infrastructure to connect these cities.

Combining reductions to mobility costs to top tier cities, with relocating pollution, we get a larger effect on GDP. GDP rises by 11.60%, and internal migration alone raises GDP by 8.07% (row 6). In contrast, the direct health and productivity effects only raise GDP by 1.60%.

Table 10: Distributional Effects of Relocating Pollution Across Cities

	Skilled Wage (%)			Unskilled Wage (%)		
	Overall	Health	Relocate+Agglom	Overall	Health	Relocate+Agglom
Relocate steady state PM2.5	13.600	2.760	9.352	0.105	0.990	-1.613
Relocate exogenous part of PM2.5	12.868	2.613	8.702	-0.742	0.856	-2.342
Relax hukou	5.154	0.000	5.154	2.661	0.000	2.661
Relax overall mobility constraints	10.025	0.000	10.025	5.368	0.000	5.368
Relocate PM2.5 & relax hukou	17.672	2.760	13.394	2.655	0.990	0.824
Relocate PM2.5 & lower mig costs	23.445	2.760	18.773	5.543	0.990	3.483

Notes: In this counterfactual exercise we relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relocate only the exogenous component of pollution. In addition to such relocations of pollution, we also relax the *hukou* restriction in the 35 high tier cities to be the same level as the median city in the country (row 3). In row 4 we relax overall migration costs to the 35 high tier cities to be the same as the median city. Columns 1-3 show the effects on skilled workers, while columns 4-6 show the effects on unskilled workers.

In Table 10 we look at the consequences for wages of skilled and unskilled workers. Relocating pollution to cities with less skill-biased capital raises skilled wages, but has little effect on the wage of the unskilled. This is a consequence of the baseline distribution of skill groups across cities that see pollution changes. As skilled workers leave cities which receive pollution, skilled wages rise in those cities and (complementary) unskilled wages fall. As skilled workers relocate to cities with more skill-biased capital, their overall productivity increases when matched with more such capital. Native skilled workers in receiving cities may see a dampening of their wages, but may also benefit from agglomeration economies. On net, we find that skilled wages rise by 13.6%, and the relocation channel raises skilled wages by 9.35%. The health channel also raises skilled wages as skilled workers already tended to locate in high skill-biased capital cities which now saw a reduction in pollution. Wages for unskilled workers barely change as the health and relocation channels cancel each other out.

Relaxing *hukou* restrictions and lower migration costs (rows 3 and 4) raise wages for both the skilled and unskilled, as workers can match better with where their marginal products are higher. That is, workers located to where there is more capital, and where there are more complementary workers. As such, a combination of lower migration costs and relocating pollution (row 6) can raise skilled wages by as much as 23.45% and unskilled wages by 5.54%, almost entirely due to changes in internal migration.

Table 11: Welfare Effects of Relocating Pollution Across Cities

	Skilled Welfare	Unskilled Welfare	Average Welfare
Relocate steady state PM2.5	26.558	1.630	3.785
Relocate exogenous part of PM2.5	26.060	1.940	4.043
Relax hukou	15.111	0.207	1.496
Relax overall mobility constraints	17.439	10.057	10.696
Relocate PM2.5 & relax hukou	47.231	1.855	5.778
Relocate PM2.5 & lower migration costs	51.479	12.415	15.793

Notes: In this counterfactual exercise we relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relocate only the exogenous component of pollution. In addition to such relocations of pollution, we also relax the *hukou* restriction in the 35 high tier cities to be the same level as the median city in the country (row 3). In row 4 we relax overall migration costs to the 35 high tier cities to be the same as the median city.

Changes to wages, however, do not capture the entirety of the welfare consequences, as pollution and migration costs also directly determine welfare. This is particularly important to acknowledge, as relocating pollution to less productive areas may be undesirable from an environmental justice point of view if it makes unskilled workers in poor cities worse off.

Table 11 examines changes to welfare by skill group. Relocating pollution away from skill-biased capital cities raises the welfare of skilled workers by 26.56%, as it raises their wages, but also lowers their experience of pollution since most skilled workers are already located in skill-biased capital cities. Unskilled workers, however, see only modest improvements in their overall welfare. Welfare in the country improves by 3.79%.

Lowering migration costs (row 4), on the other hand, raises the welfare of both skilled and unskilled workers, by 17.44% and 10.06% respectively. The combination of relocating pollution and lowering migration costs improves welfare in the country by 15.79%.

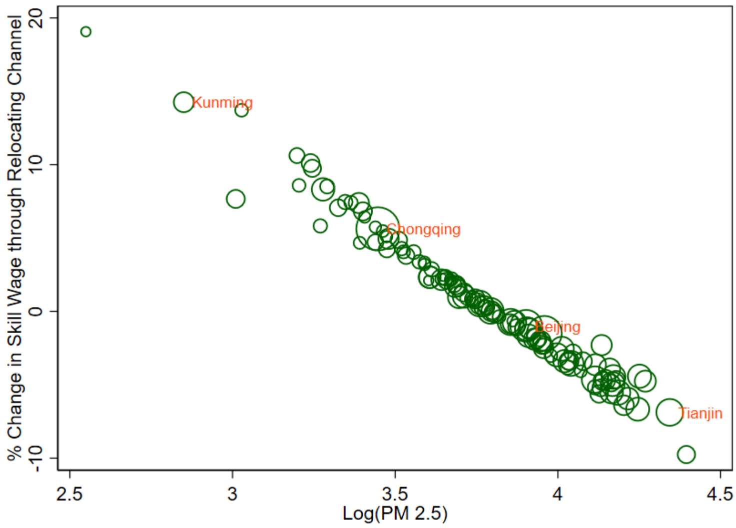
Once again, in Appendix Table D2, we re-examine the overall changes to GDP from our most basic formulation of the model, without externalities. That is, without housing, agglomeration, or pollution responses to changing populations. The results are qualitatively similar to Table 9, with some improvements to GDP being slightly larger in magnitude.

8.3 How Much of the Cross-city Wage Gap is Due to Pollution?

Next, we return to our initial motivation in explaining, why despite existent productivity gaps across regions, worker mobility does not equalize wages. That is, how much of the cross-city wage gaps can be explained by poor air quality.

We conduct an exercise where we change the amount of pollution in all cities to be that of the median city in the country, while still keeping the average level of pollution the same as before. This means raising pollution levels in low-polluting low skill-biased capital cities, and lowering them in high-polluting productive cities. This exercise allow us to quantify how much of the wage gap across cities is because of pollution-induced relocation. In Figure 12 we show how wages would change with this reallocation of pollution. As pollution is lowered in cities like Beijing and Tianjin, 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 Kunming is bridged by 33.68%, suggesting that pollution determines about one-third of the wage gaps between these two cities.

Figure 12: Explaining the Wage Gap



Notes: We plot the change in the skilled wage, solely due to changes in worker location, 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.

8.4 Consequences of the 2013 City-level Pollution Caps

Finally, we quantify the productivity implication of China's recent pollution control policy. On Sep 10, 2013, State Council of China issued *Air Pollution Prevention and Control Plan*. The specific pollution control target was stated to be, "by 2017, annual PM2.5 concentration in China's three major economic circles: Beijing-Tianjin-Hebei, Yangtze River Delta and Pearl River Delta region shall respectively fall by around 25%, 20% and 15%. PM2.5 concentration in Beijing shall be controlled below 60 $\mu g/m^3$." This plan mainly targeted China's most productive areas, which would induce skilled workers to resort into those areas.

Table 12: The Productivity Effect of Pollution Regulation

	Change in GDP per Worker (%)			
	Overall changes	Health	Relocation	Relocate+Agglom
Control PM2.5	2.677	0.947	1.462	1.668
Control PM2.5 & relax hukou	5.351	0.947	3.815	4.282
Control PM2.5 & lower migration costs	9.784	0.947	7.904	8.672

Notes: In this counterfactual exercise we reduce pollution according to the targets set by the 2013 *Air Pollution Prevention and Control Plan* (row 1). In addition to pollution regulations, we also relax the hukou restriction in higher tier cities (row 2), and migration costs (row 3) by 50%. Column 1 shows the gain to overall GDP. Column 2 shows the gain to GDP stemming from the health-productivity channel. Column 3 shows the GDP change from the relocation channel, and Column 4 also accounts for changes in TFP due to changes in the agglomeration of skilled workers.

As shown in Table 12, the recent air pollution control policy, targeted at only a handful of cities, would increase country-level GDP per worker in China by 2.67%, mostly driven by workers relocating. When combined with lowering migration costs GDP per worker would increase by 9.78%. The relocation channel alone raises GDP per worker by 7.9%. As such, overlooking the productivity effect via migration would substantially understate the economic benefit of pollution control policy.

Table 13: Distributional Effects of Pollution Regulation

	Skilled Wage			Unskilled Wage		
	Overall	Health	Relocate+Agglom	Overall	Health	Relocate+Agglom
Control PM2.5	2.653	0.926	1.658	2.689	0.957	1.673
Control PM2.5 & relax hukou	6.648	0.926	5.571	4.687	0.957	3.623
Control PM2.5 & lower mig costs	12.957	0.926	11.814	8.161	0.957	7.065

Notes: In this counterfactual exercise we reduce pollution according to the targets set by the 2013 *Air Pollution Prevention and Control Plan* (row 1). In addition to pollution regulations, we also relax the hukou restriction in higher tier cities (row 2), and migration costs (row 3) by 50%. Columns 1-3 show effects on skilled wages. Columns 4-6 on unskilled wages.

We again examine how these policy changes affect wages for the different types of workers in Table 13. The pollution caps, again targeted at only a few cities, combined with lower migration costs, raise skilled wages by 12.96%, and unskilled wages by 8.16% in the country. Finally, in Table 14, we examine the welfare consequences of these pollution caps. The control itself raises the welfare of the skilled by 6.6%, but barely changes the welfare of the unskilled. Combined with lower migration costs, though, both skilled and unskilled welfare rise, and combined welfare increases by 12.18%.

Table 14: Welfare Effects of Pollution Regulation

	Skilled Welfare	Unskilled Welfare	Average Welfare
Control PM2.5	6.606	0.767	1.272
Control PM2.5 & relax hukou	17.764	1.662	3.054
Control PM2.5 & lower migration costs	25.941	10.872	12.175

Notes: In this counterfactual exercise we reduce pollution according to the targets set by the 2013 *Air Pollution Prevention and Control Plan* (row 1). In addition to pollution regulations, we also relax the hukou restriction in higher tier cities (row 2), and migration costs (row 3) by 50%.

9 Conclusion

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 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

Adao, R., M. Kolesar, and E. Morales (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics 134*(4), 1949–2010.

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

Ahlfeldt, G., S. Redding, D. Sturm, and N. Wolf (2015). The Economics of Density: Evidence from the Berlin Wall. *Econometrica 83*(6).

Allen, T., C. Arkolakis, and Y. Takahashi (2020). Universal Gravity. *Jounral of Political Economy 128*(2), 393–433.

Amiti, M. and C. Freund (2010, March). *The Anatomy of China's Export Growth*, pp. 35–56. University of Chicago Press.

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.

Autor, D., D. Dorn, and G. H. Hanson (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review 103*(6), 2121–68.

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.

Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-experimental shift-share research designs. Technical report, National Bureau of Economic Research.

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. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

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.

Giles, J., A. Park, and M. Wang (2019, October). The Great Proletarian Cultural Revolution, Disruptions to Education, and the Returns to Schooling in Urban China. *Economic Development and Cultural Change 68*(1).

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. Technical report, American Economic Review. Forthcoming.

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*.

Jaeger, D. A., J. Ruist, and J. Stuhler (2018). Shift-share instruments and the impact of immigration. *National Bureau of Economic Research, WP 24285*.

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

Jia, Barwick, P., S. Li, L. Lin, and E. Zou (2019). From fog to smog: The value of pollution information. *National Bureau of Economic Research, WP 26541*.

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

Khanna, G., K. Shih, A. Weinberger, M. Xu, and M. Yu (2020). Trade Liberalization and Chinese Students in US Higher Education. *Working Paper*.

Kinnan, C., S.-Y. Wang, and Y. Wang (2018). Access to Migration for Rural Households. *American Economic Journal: Applied Economics 10*(4), 79–119.

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.

Lustgarten, A. (2020, September). How Climate Migration Will Reshape America. *New York Times*.

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, November). 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 wide-range of meaningful robustness checks to evaluate the concreteness of the empirical relationship between air quality and migration decisions. We explore threats to identification, different instrumental variables, alternative model specifications and data sources.

A.1 Alternative Instruments and Sources of Variation

First, we explore a few different sources of underlying variation. We employ two different instrumental variables strategies discussed in recent work on China to address the endogeneity of air pollution. Specifically, we study the variation underlying the instrument based on wind direction and distant 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. 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 thus exclude any plants built within different distance radii around the city, and find a similar empirical pattern. 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 show that the IV is not predicted by baseline city-level characteristics.

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.

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 by skill.

A.1.1 Instruments Variable Estimates

In our main text, we use our first instrument based on wind direction and distant coal-fired plants to deal with the endogeneity of local air pollution. We show how in Table 1 there is a strong relationship between poor air quality and the out-migration of high-skilled workers. In this section, we employ an alternative instrumental strategy. We isolate exogenous fluctuations in air pollution by leveraging the variation in thermal inversions that trap pollutants. In Table A1 we show the strength of the first stage relationships between our different instruments and our independent variable of interest.

In Table A2, we study how variation in PM 2.5 from our second instrument of thermal inversions affects migration tendency. We include distances to three large seaports to account for the spatial distribution of economic development in China. In the first three columns, we employ the annual occurrence of thermal inversions as the instrumental variable. The results show that the implications of air pollution on emigration are more pronounced for high-skilled workers in comparison with those with lower skills. A 10% increase in PM 2.5 leads to a 1.79 percentage point increase in out-migration rates for those with high education attainment, but only 1.08 for those without. In the last three column, we leverage the variation coming from the annual strength of thermal inversions. The results remain similar.

Since thermal inversions may be affected by local weather conditions, we further control for a set of weather characteristics in Table A3. We find a similar empirical pattern–high-skilled

Table A1: The First Stage Across Different Instruments

	Dependent variable: Log (PM2.5)					
Wind Direction and Coal Plants	0.00837***			0.00846***		
	(0.00168)			(0.00166)		
Number of inversions		0.00110***			0.00152***	
		(0.000260)			(0.000230)	
Strength of inversions			0.000318**			0.000541***
			(0.000145)			(0.000140)
Observations	285	285	285	285	285	285
F-value	47.4	32.89	21.49	27.99	24.58	17.59
Adjusted R-squared	0.398	0.413	0.375	0.442	0.495	0.450
City controls	Y	Y	Y	Y	Y	Y
Weather controls	N	N	N	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 and to Tianjin seaport and to Shenzhen seaport. Weather controls include termperature, wind speed, sunshine duration and humidity.

workers are more likely to escape from polluted cites to avoid the adverse effects of pollution, compared to their lower educated counterparts.

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

While our earlier results examine the out-migration decision of individuals, we further examine the association between air pollution and destination choices of workers. We regress city-level in-migration rate on local air pollution concentration. The results are shown in Table A5 and A6. The number of cities in our sample decreases a little to 273, because we drop the cities that skill-specific in-migration rate is missing. We use wind direction and coal-fired plants IV to deal with the endogeneity of air quality in destination cities in Tables A5, and employ thermal inversions IV in Table A6. We find that even for in-migration decisions, the response of high-skilled workers is greater than that of low-skilled workers. In other words, high-skilled workers are more likely to move to cities with clean air when they make location choices. Severe air pollution can not only result in the outflow of high-skilled workers but also reduce their inflow, which in-turn lower the skill ratio of workers.

Table A2: Thermal Inversions IV

	Dependent variable: Leave hukou city indicator					
Instruments:	Number of inversions			Strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.113***	0.108***	0.179***	0.0790*	0.0733	0.189***
	(0.0373)	(0.0412)	(0.0480)	(0.0424)	(0.0465)	(0.0671)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.027	0.026	0.066	0.030	0.029	0.064
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 and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.Weather controls include termperature, wind speed, sunshine duration and humidity.

Table A3: Thermal Inversions IV: Controlling for Weather Conditions

	Dependent variable: Leave hukou city indicator					
Instruments:	Number of inversions			Strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0993***	0.0951***	0.133***	0.0781***	0.0727**	0.126***
	(0.0276)	(0.0295)	(0.0313)	(0.0302)	(0.0321)	(0.0363)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.041	0.041	0.078	0.043	0.044	0.079
City controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y
Weather controls	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 and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.Weather controls include termperature, wind speed, sunshine duration and humidity.

Table A4: Combined Instruments

	Dependent variable: Leave hukou city indicator					
Instruments:	Wind and Coal + Number of Inversions			Wind and Coal + Strength of Inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log(PM 2.5)	0.0940***	0.0877***	0.127***	0.0801***	0.0717**	0.122***
	(0.0277)	(0.0301)	(0.0264)	(0.0294)	(0.0319)	(0.0279)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.041	0.042	0.079	0.043	0.044	0.080
Hansen J statistic	0.276	0.535	0.101	0.018	0.005	0.026
Hansen P value	0.5995	0.4646	0.7511	0.8919	0.9462	0.8712
City controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y
Weather controls	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, and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. Weather controls include termperature, wind speed, sunshine duration and humidity.

Table A5: Pollution and In-migration: Wind Direction and Coal-fired Plants IV

	Dependent variable: Share of in-migrants		
	Full Sample	Low edu	High edu
Log (PM 2.5)	-1.113*	-1.101	-1.655**
	(0.653)	(0.686)	(0.756)
Observations	273	273	273
Adjusted R-squared	0.148	0.168	-0.044
City controls	Y	Y	Y

Notes: Dependent variable is the log share of in-migrants in city population. Independent variable is destination city PM 2.5. Standard errors clustered at the 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 A6: Pollution and In-migration: Thermal Inversions IV

	Dependent variable: Share of in-migrants					
Instruments:	Number of inversions			Strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
Log (PM 2.5)	-1.143***	-1.069**	-1.746***	-1.119**	-0.969*	-2.183***
	(0.427)	(0.427)	(0.557)	(0.504)	(0.502)	(0.740)
Observations	273	273	273	273	273	273
Adjusted R-squared	0.143	0.168	-0.060	0.146	0.176	-0.167
City controls	Y	Y	Y	Y	Y	Y
Weather controls	Y	Y	Y	Y	Y	Y

Dependent variable is the log share of in-migrants in city population. Independent variable is destination city PM 2.5. Standard errors clustered at the city level are reported in parentheses. Instrumental variables specification using the number of thermal inversions and the strength of thermal inversions. 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.

A.1.2 Endogeneity Concerns over Instrumental Strategies

In this section we test the identification assumptions underlying our two instruments. Our first instrument is the interaction between the three components of wind direction, distance and coal consumption. We may expect that policy makers take these components into account when placing large coal-fired plants near certain type of cities. For example, those plants may be systematically built near poorer, less influential cities. Thus, there is a concern that the instrument may be correlated with unobservable characteristics of nearby cities. In Table A7, we exclude any plants built within 200km of a given city (first two columns), and then within 400km of the city (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)	0.0895*	0.185***	0.113***	0.204***
	(0.0497)	(0.0474)	(0.0418)	(0.0493)
Observations	368,957	75,533	368,957	75,533
Adjusted R-squared	0.028	0.064	0.026	0.060
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 and to Shenzhen seaport. 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 construction of IV, and instead include cities with newly built ones in the 'control' group. Although this empirical strategy is more conservative, we find similar patterns and magnitudes.

We may also expect that coal-fired plants plants are more likely to be located in coal producing areas. In as far as coal producing areas may also influence the underlying industrial structure, this may raise concerns about other unobservable associations with migration rates. Shanxi is the largest coal producing province in China, we thus exclude Shanxi in our estimation. As reported in In Table A9 , the results are slightly more precisely estimated.

Another concern is that power plants may be built near cities that require more electricity. Even though those plants supply electricity to vast areas of China including many remote provinces (Freeman, Liang, Song, and Timmins, 2019), we examine this concern by controlling for city-level electricity consumption. We include total electricity consumption in the first two columns in Table A10, and add industrial electricity consumption in the last two columns. Accounting for city-level electricity consumption hardly change our empirical pattern.

Our wind-direction and coal-fired plants IV is constructed based on the interaction between wind direction, distance to coal plant and coal consumption at power plant. It is natural to ask which of the three components drives our IV results. In Table A11 we try different versions of the

Table A8: Excluding Newly Built Power Plants

	Dependent Variable: Leave hukou city indicator							
	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: Cities 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: Without Coal Producing Province

Dependent variable: Leave hukou city indicator		
	Low edu	High edu
Log(PM 2.5)	0.0548	0.114***
	(0.0502)	(0.0390)
Observations	350,995	72,068
Adjusted R-squared	0.028	0.076
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 and to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A10: Controlling for Local Electricity Consumption

	Dependent variable: Leave hukou city indicator			
Additional controls:	Total Electricity Consumption		Industrial Electricity Consumption	
	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.00709	0.0793**	0.0120	0.0899***
	(0.0560)	(0.0333)	(0.0566)	(0.0335)
Observations	357,333	73,943	357,333	73,943
Adjusted R-squared	0.031	0.089	0.031	0.089
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 and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

instrument in the construction of which we exclude each of the three components, respectively. Our results hardly change when we exclude distance to coal plant and coal consumption at power plant. In contrast, the coefficient estimates of air pollution become statistically insignificant when we exclude the component of wind direction. Therefore, our main IV results are mainly driven by the variation of wind direction across locations. Wind direction is determined by nature and remains stable over long periods of time, thus it can be consider as exogenous to local economic attributes.

Table A11: Decomposing the Wind Direction IV

	Dependent variable: Leave hukou city indicator					
	IV:Excluding distance		IV:Excluding coal consumption		IV:Excluding wind direction	
	Low edu	High edu	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0908	0.172***	0.0505	0.111***	-0.0377	0.0411
	(0.0589)	(0.0464)	(0.0512)	(0.0389)	(0.0531)	(0.0387)
Observations	368,957	75,533	368,791	75,467	368,957	75,533
Adjusted R-squared	0.028	0.067	0.029	0.075	0.023	0.077
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), the interaction between wind direction and distance to plant (next two columns), and the interaction between distance to plant and coal consumption (last two columns). 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.

One concern with the exogeneity of wind direction is that the Chinese government might select the thermal plants in a way that pollution did not travel to the populated cities or politically important cities. If that were the case, coal-fired plants are less likely to be located upwind of such cities. In Table A12, we present the number of large-scale thermal plants located up wind of five largest metropolitan areas in 2014 along with their total coal consumption. Among them, Beijing and Tianjin are most populated and politically important cities in Northern China, but highly polluted. There are 58 large scale coal-fired plants located upwind of Beijing, and 59 large scale thermal power plants located upwind of Tianjin. The total consumption of them is as large as 2372.747 million tons and 2123.982 million tons for Beijing and Tianjin, respectively. The coal consumption of large coal-fired plants is higher for Beijing and Tianjin compared to the mean of that for all Chinese cities, indicating that more households suffer from severe air pollution in the two large cities. We also calculate the ratio of the upwind large thermal plants to the total number of large thermal plants. As presented in column 2, the "ratio" is 22.4% and 50.8% for Beijing and Tianjin, respectively. The mean of the "ratio" for all Chinese cities are 37.8%, which is between the values in Beijing and Tianjin. Thus, Chinese government do not locate the coal-fired plants in a way that wind do not blow pollutants to the populated cities or politically important cities.

Table A12: The Coal-fired Plants Located Upwind of Large metropolitans

City	Number of Upwind Plants	Ratio of Upwind Plants	Coal Consumption of Upwind Plants	Smallest Angle of Plants
Beijing	58	22.414%	2372.747	25.018
Tianjin	59	50.847%	2123.982	26.608
Shanghai	62	1.613%	2067.868	16.028
Guangzhou	30	16.667%	899.35	22.233
Shenzhen	25	44.000%	785.098	26.58
National mean	43.157	37.818%	511.597	25.897

Notes: The statistics are calculated using the large-scale thermal power plants located outside a given city and within 500km. Following (Freeman, Liang, Song, and Timmins, 2019), we define upwind region as a section of a circular buffer drawn at a distance of 500 km from a given city, and the angle between the left/right side of the section and the annual dominant wind direction of the city is 45 degree.

To further test whether politicians avoid populated, politically important and rich 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 A13, we explore whether city-level characteristics in 2005 can predict (a) the ratio of upwind plants built after 2005, and (b) the IV based on plants built after 2005. We find no meaningful associations between these variables and possible predictors of a city's influence, like baseline populations, GDP, total 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).

In Table A14, we construct placebo instruments based on artificially changing the wind direction angle by first 90 degrees, and then 180 degrees. The first two columns report the first stage results. As the angle is increased, the falsified instrument is less likely to predict PM 2.5. The last four columns show that we can not find similar empirical pattern when we employ the falsified IV.

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 the number and the strength of thermal inversions in Table A15. We fail to find any such meaningful associations. Furthermore, we also find that lagged inversions do not predict future inversions, suggesting that the levels

Table A13: Baseline Economy and the Wind Direction IV

Dependent variable:	The ratio of upwind plants		Wind direction and Coal plants IV	
Baseline Population	-0.009	-0.013	-93.886	-103.955
	(0.027)	(0.025)	(138.977)	(130.149)
Baseline GDP per capita	0.020	0.013	-165.748	-202.833
	(0.040)	(0.038)	(240.623)	(221.615)
Baseline Elec cons	-0.019		14.687	
	(0.021)		(97.437)	
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 2005, independent variables are measured in the year 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 and to Shenzhen seaport.

Table A14: Placebo Wind Directions

Dependent variable:	Log(PM 2.5)		Leave hukou city indicator			
			Low edu	High edu	Low edu	High edu
Log(PM 2.5)			-0.121	0.0259	-0.156	-0.0130
			(0.0948)	(0.0680)	(0.106)	(0.0696)
Coal IV Placebo	0.00524*					
(wind direction plus 90 degrees)	(0.00284)					
Coal IV Placebo		0.00724				
(wind direction plus 180 degrees)		(0.00501)				
Observations	285	285	368,957	75,533	368,957	75,533
Adjusted R-squared	0.369	0.366	0.004	0.077	None	0.074
City controls	Y	Y	Y	Y	Y	Y
Demographics	N	N	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, and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

of auto-correlation in inversions are low, and we may consider the data generating process underlying inversions to be close to random.

Table A15: Lagged Pollution and Thermal Inversions

Dependent variable:	Number of inversions		Strength of nversions	
Lagged Log(PM 2.5)	-1.603	-1.602	-3.632	-3.888
	(2.057)	(2.054)	(4.971)	(4.992)
Lagged number of inversions		-0.0269		
		(0.0200)		
Lagged 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
City fixed effects	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y

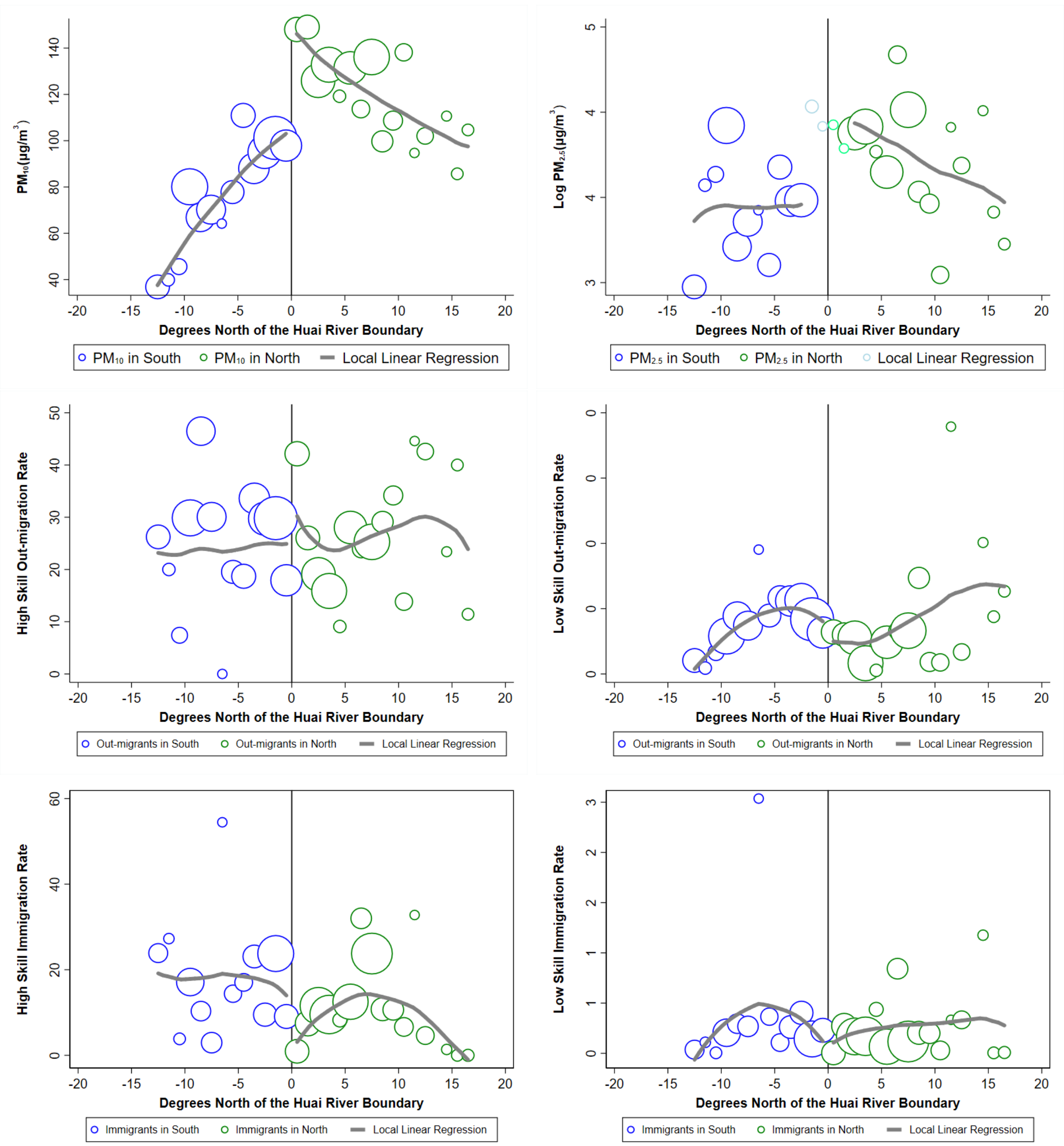
Notes: Standard errors clustered at city level are reported in parentheses. City level regressions for 337 cities over 16 years. Specifications include city and year fixed effects.

A.1.3 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 A1 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 A1: 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 specifications, sample restrictions and control variables. Once again, our aim is to test the robustness of our estimates to various empirical concerns.

First, we employ an individual-level longitudinal panel data to estimate the association between air quality and workers' spatial sorting. The longitudinal data allow us to track individuals over time and control for both individual- and city-level unobservables. Importantly, we use an alternative definition for migration status regardless of *hukou* location. The estimated effects of pollution on the out-migration of high-skilled workers increase in magnitude when we use the alternative definition for migration status.

Then, we turn to the implications of cumulative pollution. We find that workers are more sensitive to cumulative pollution when they make location choices, compared to short-term pollution. Once again, the impacts of cumulative pollution are more pronounced for high-skilled workers than their low-skilled counterparts.

In the section that follows thereafter, we use alternative samples to examine the effects of pollution on migration decisions. Using different samples does little to affect our empirical pattern. Our results are also robust to employing an alternative measure of local air quality.

Finally, we add a wide range of covariates that may confound the relationship between air quality and internal migration. Incorporating additional controls hardly change our results.

A.2.1 Individual Longitudinal Panel and Alternative Definition of Migration

We employ an individual-level longitudinal panel and a different definition of migration status to explore the spatial sorting of Chinese workers. We use an alternative data source, the China Labor-force Dynamic Survey (CLDS), which is a national social survey targeted at labor force dynamics in China. CLDS 2014 asks a retrospective history of locations for individuals. We use this retrospective location history to create an individual-level longitudinal panel between 2006 and 2014. Here, we define migration to be an indicator for whether or not an individual changed city location between years, regardless of whether they change their *hukou* location or not. The strengths of the individual-level panel lie in that it allows us to account for individual-specific unobservables and track those who have moved multiple times and who have moved and returned home. We combine the longitudinal data with annual information on PM 2.5 at the city level.

Since data on large-scale coal-fired plants are not available for each year during the sample period, we employ the instrument of the number of thermal inversions to address the endogeneity of air pollution. Table A16 presents the IV estimation of the relationship between air pollution and the out-migration tendency of individuals, controlling for city-, year- and individual-fixed effects. Including individual-specific fixed effect allows us to account for individual-level unobservables (such as tastes for clean air, individual preference for a specific city) that may be systematically correlated with migration decisions. The results show that once again, the response of high skill migration to air pollution is a lot stronger than that of low-skill workers. A 10% increase in PM 2.5 raises out-migration rates by 1.39 percentage points for high-skilled workers. In our main text, migrants are defined as those who are away from their *hukou* city. Under this definition, we may miss those who move to a different city and obtain local *hukou* in the city. Thus, we may understate high-skilled workers' migration response to pollution exposure, because migrants with high education attainment are much easier to obtain local *hukou* than those without. In line with our expectation, the effects on high-skill migration shown in Table A16 are larger in magnitude than our baseline estimates, because we consider both non-*hukou* migration (change residential location without changing *hukou* location) and *hukou* migration (change both residential location and *hukou* location) in the definition of migration status.

One concern with the thermal inversions IV is that it may be affected by weather conditions. We thus control for weather amenities in Table A17. The coefficients estimates are quantitatively and qualitatively similar.

Table A16: Individual Longitudinal Panel

Dependent variable:	Change city location indicator		
	Full sample	Low edu	High edu
Log (PM2.5)	0.0305	0.00520	0.136**
	(0.0237)	(0.0245)	(0.0680)
Observations	123,801	105,502	18,297
Adjusted R-squared	0.211	0.220	0.171
City fixed effect	Y	Y	Y
Year fixed effect	Y	Y	Y
Individual fixed effect	Y	Y	Y
Demographics	Y	Y	Y

Notes: Standard errors clustered at the individual level are reported in parentheses. Instrumental variables specification using the number of thermal inversions. Demographics include age, age-squared, marital status,an urban hukou indicator and a past migration experience indicator.

Table A17: Individual Longitudinal Panel:Controlling for Weather Amenities

Dependent variable:	Change city location indicator		
	Full sample	Low edu	High edu
Log (PM2.5)	0.0301	0.00424	0.139**
	(0.0247)	(0.0256)	(0.0700)
Observations	123,801	105,502	18,297
Adjusted R-squared	0.211	0.220	0.170
City fixed effect	Y	Y	Y
Year fixed effect	Y	Y	Y
Individual fixed effect	Y	Y	Y
Demographics	Y	Y	Y
Weather controls	Y	Y	Y

Notes: Standard errors clustered at the individual level are reported in parentheses. Instrumental variables specification using the number of thermal inversions. Demographics include age, age-squared, marital status,an urban hukou indicator and a past migration experience indicator. Weather controls include temperature, wind speed, sunshine duration and humidity.

In China, migrant workers are more likely to move from the inland region to the coastal region to gain access to better economic opportunities. As a result, the vast majority of migrants are concentrated in large coastal cities. To account for the potential role played by differential

Table A18: Individual Longitudinal Panel: Controlling for Region-Specific Trend

Dependent variable: Change city location indicator	Full sample	Low edu	High edu
Log (PM2.5)	0.0434	0.0167	0.156**
	(0.0277)	(0.0289)	(0.0769)
Observations	123,801	105,502	18,297
Adjusted R-squared	0.210	0.220	0.167
City fixed effect	Y	Y	Y
Year fixed effect	Y	Y	Y
Individual fixed effect	Y	Y	Y
Demographics	Y	Y	Y
Weather controls	Y	Y	Y
Region trend	Y	Y	Y

Notes: Standard errors clustered at the individual level are reported in parentheses. Instrumental variables specification using the number of thermal inversions. Demographics include age, age-squared, marital status,an urban hukou indicator and a past migration experience indicator. Weather controls include temperature, wind speed, sunshine duration and humidity.

migration pattern between coastal and inland China, we further add region-specific trend in A18. Including region-specific trend does little to affect our estimates.

A.2.2 Accumulated Pollution over Time

As migration decisions are long-lasting, we expect that people are more likely to response to accumulated pollution, compared to contemporaneous pollution shocks. While we measure out-migration in 2015, we wish to understand how migration decisions depend on the cumulative PM 2.5 concentration over different time intervals. Since early data on large-scaled coal-fired power plants are not available, we use the occurrence of thermal inversions averaged over different time periods to deal with the endogeneity of cumulative air pollution.

In Table A19, we use specifications where PM2.5 are averaged over 5, 10, 15 and 18 years, respectively, leading up to the migration decision. We find that the longer the time period of PM 2.5 exposure, the larger is the response. As such, pollution exposure in a short time frame have a smaller impact than the same amount of exposure spread out over a longer time period. Once again, we see similar empirical pattern as our baseline estimates. The effects of cumulative pollution on emigration are also more pronounced for skilled workers, relative to their unskilled counterparts.

In Table A20, we further add cumulative weather conditions that may confound the association between the thermal inversions IV and migration decisions. We consistently find stronger effects on out-migration for the higher educated group of workers.

Table A19: 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
Log (Mean PM2.5: 1998-2015)	0.146***	0.141***	0.241***			
	(0.0483)	(0.0537)	(0.0681)			
Log (Mean PM2.5: 2001-2015)				0.140***	0.134***	0.231***
				(0.0460)	(0.0510)	(0.0647)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.023	0.023	0.051	0.023	0.023	0.053
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
Log (Mean PM2.5: 2006-2015)	0.129***	0.122***	0.218***			
	(0.0424)	(0.0470)	(0.0608)			
Log (Mean PM2.5: 2011-2015)				0.116***	0.109**	0.205***
				(0.0402)	(0.0448)	(0.0570)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.025	0.024	0.056	0.026	0.025	0.060
City controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Instrumental variables using the number of thermal inversions averaged 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 and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A20: PM 2.5 Measured over Different Time Intervals: Controlling for Weather Conditions

	Dependent variable: Leave hukou city indicator					
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
Log (Mean PM2.5: 1998-2015)	0.130***	0.119***	0.208***			
	(0.0331)	(0.0338)	(0.0559)			
Log (Mean PM2.5: 2001-2015)				0.120***	0.108***	0.201***
				(0.0316)	(0.0325)	(0.0546)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.037	0.041	0.061	0.038	0.041	0.062
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y
Weather Controls	Y	Y	Y	Y	Y	Y

	Dependent variable: Leave hukou city indicator					
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
Log (Mean PM2.5: 2006-2015)	0.144***	0.135***	0.205***			
	(0.0367)	(0.0392)	(0.0503)			
Log (Mean PM2.5: 2011-2015)				0.105***	0.0960***	0.165***
				(0.0323)	(0.0350)	(0.0441)
Observations	444,490	368,957	75,533	444,490	368,957	75,533
Adjusted R-squared	0.028	0.029	0.061	0.037	0.037	0.072
City Controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y
Weather Controls	Y	Y	Y	Y	Y	Y

Notes: Instrumental variables using the number of thermal inversions averaged 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 and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. Weather controls include temprature, humidity, sunshine duration and wind speed averaged over different time intervals.

A.2.3 Alternative Samples and Alternative Measure of Air Quality

In this section, we use different samples to explore the relationship between pollution and out-migration. Then, we use an alternative measurement of air quality to examine our empirical pattern.

First, we look at different ways to slice the data. In Table A21, instead of splitting up the sample into low and high skilled, we split it up into three categories: high school or bellow, those with some college education, and those with college or above education. 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 A21: Disaggregated Education Levels

	Dependent variable: Leave hukou city indicator		
	High school or below	Some college	College or above
Log (PM 2.5)	0.0302	0.0729*	0.150***
	(0.0596)	(0.0438)	(0.0504)
Observations	368,957	39,287	36,246
Adjusted R-squared	0.029	0.075	0.080
City controls	Y	Y	Y
Demographics	Y	Y	Y

Notes: Standard errors clustered at the individual 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.

We next 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.

We further conduct heterogeneity analysis. In Table A23 ,we split up the sample into three age groups. We see that the implications of pollution on emigration are stronger for younger workers. There are two reasons for the differential migration pattern by age. First, young people have less mobility costs compared to their older counterparts. Second, they have better have better knowledge about the adverse effects of air pollution.

In Table A24 we split up the sample by rural and urban origin locations. We find that the elasticity for high-skill out-migration is larger in rural areas than it is in urban areas.

Finally, we turn our attention to an alternative measure of local air quality. 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. Recall that Air Quality Index (AQI) is an overall indicator for air pollution concentration calculated using multiple atmospheric pollutants including SO2, NO2, PM10, PM2.5, O3 and CO. As shown in Table A25, our empirical pattern are robust to using the AQI as our independent variable of interest.

A.2.4 Additional Controls

In this section, we include various sets of controls that may confound the association between local air pollution and migration decisions. In the first two columns of Table A26, we account

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

	Dependent variable: Leave hukou city indicator					
	Exclude Beijing		Exclude Tianjin		Exclude Shijiazhuang	
	Low edu	High edu	Low edu	High edu	Low edu	High edu
Log (PM 2.5)	0.0303	0.103**	0.124	0.142**	0.0315	0.104**
	(0.0598)	(0.0432)	(0.0805)	(0.0551)	(0.0599)	(0.0440)
Observations	366,760	72,655	363,518	72,261	367,439	75,096
Adjusted R-squared	0.029	0.078	0.029	0.071	0.029	0.076
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)	0.0294	0.102**	0.0324	0.116***	0.0304	0.105**
	(0.0594)	(0.0428)	(0.0606)	(0.0443)	(0.0600)	(0.0443)
Observations	367,827	74,884	366,287	74,358	367,741	74,736
Adjusted R-squared	0.029	0.076	0.029	0.076	0.029	0.076
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 and to Shenzhen seaport. 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)	0.00689	0.156**	0.0357	0.0600*	0.0388	0.0500**
	(0.0839)	(0.0677)	(0.0648)	(0.0309)	(0.0445)	(0.0206)
Observations	105,397	38,304	123,326	23,457	140,234	13,772
Adjusted R-squared	0.019	0.059	0.020	0.062	0.013	0.013
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 and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A24: 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)	0.0198	0.0951***	0.0486	0.267**
	(0.0409)	(0.0359)	(0.0884)	(0.115)
Observations	165,436	65,911	203,521	9,622
Adjusted R-squared	0.024	0.028	0.033	0.045
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 and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A25: Air Quality Index

	Full sample	Low edu	High edu
Log (AQI)	0.0722	0.0384	0.150**
	(0.0742)	(0.0811)	(0.0707)
Observations	441,943	366,717	75,226
Adjusted R-squared	0.026	0.027	0.069
City Controls	Y	Y	Y
Demographics	Y	Y	Y

Notes: Independent variable is Log (Annual 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, and to Shenzhen seaport. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

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 skill-specific migration. We find our estimates display similar patterns as before if we add controls for the teacher-student ratio, the number of hospitals and doctors per capita in 2000.

In the third and fourth columns of Table A26, we include controls for local economic production in 2000. Our economic controls include baseline measures of GDP per capita, as well as the size of the services and manufacturing sector. Our results are not meaningfully affected by these controls.

Fine particle concentration tends to be correlated with local industrial pollutant emissions. To account for the potential role played local industrial emissions, we add industrial SO2 emission, waste water emission and dust emission as covariates in last two columns of Table A26. The inclusion of these industrial emissions does little to affect the impacts of PM2.5 concentration.

Table A26: Additional Control Variables

	Dependent variable: Leave hukou city indicator					
Additional Controls:	Baseline skill controls		Baseline economy controls		Industrial emmisions controls	
	Low edu	High edu	Low edu	High edu	Low edu	High edu
Log(PM 2.5)	0.0379	0.0897**	0.0235	0.0890**	0.0224	0.0969**
	(0.0545)	(0.0449)	(0.0560)	(0.0451)	(0.0574)	(0.0393)
Observations	347,283	72,904	347,283	72,904	347,283	72,904
Adjusted R-squared	0.036	0.082	0.039	0.086	0.031	0.083
City controls	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

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. Baseline skill controls include teacher student ratio, log hospitals per capita and log doctors per capita in 2000. Baseline economy controls include log GDP per capita and industrial structure (the product value at service sector / manufacture sector). Industrial emissions controls include log industrial SO2 emission, log industrial waste water emission and log industrial dust emission. 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.

B Appendix: Air Pollution Data Disclosure in China

Despite the hazardous level of exposure to pollution, Chinese citizens used to have limited or no access to information about local air quality. Since 2000, there are three waves of air pollution data disclosure in China: First, in 2000, the Chinese government started to publish air quality data, including an Air Pollution Index (API) and PM10, but only did so for 42 cities. Although fine particles (i.e. PM2.5) are more hazardous than larger particles (i.e. PM10) with respect to mortality, cardiovascular and respiratory endpoints, PM2.5 was not included in the calculation of API. The number of cities in which API and PM10 were available increased gradually to 120 in 2012.

The second wave was initiated when the U.S. Embassy in China started to publish PM2.5 data. This happened first in April 2008 in Beijing, then in Shanghai and Guangzhou from near the end of 2011, and Chengdu since June 2012, and Shenyang since April, 2013.

Last, in response to the public demand for the publication of PM2.5 data, the Chinese government started to disclose real time PM2.5 data in large and median sized Chinese cities from 2012. Finally, the information on real time PM2.5 was made available in all Chinese cities since January 1, 2015.

The disclosure of pollution information has an important effect on households avoidance behaviors. As illustrated in Figure B1, the sales of indoor air filtration increases sharply in response to the information shock of PM2.5 data disclosure in 2012. Table B1 shows that PM2.5 data disclosure has significant positive impacts on the outmigration of Chinese individuals, and such impacts are more pronounced for high-skilled workers.

Figure B1: The Number of Air Purifier Sales from 2006 to 2014

Notes: Air purifier sales transaction data collected by a marketing firm in China from January 2006 through December 2014 for 85 major Chinese cities.

Table B1: PM2.5 Data Disclosure and Outmigration

Dependent variable:	Change city location indicator			Move to less polluted city indicator		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
PM2.5 data disclosure indicator	0.00609***	0.00593***	0.00680*	0.00368***	0.00343***	0.00663**
	(0.00118)	(0.00121)	(0.00363)	(0.000888)	(0.000888)	(0.00317)
City fixed effect	Y	Y	Y	Y	Y	Y
Year fixed effect	Y	Y	Y	Y	Y	Y
Individual fixed effect	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered at individual level are reported in parentheses. Demographics include age, age-squared, marital status,an urban hukou indicator and a past migration experience indicator.

C Appendix: Additional Tables and Figures

Table C1: 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 C2: 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

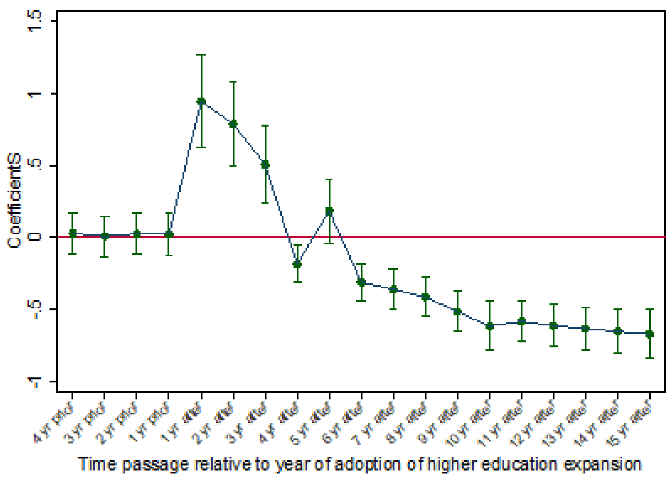
Table shows a few examples of *hukou* requirements for city workers.

Table C3: Preferences for Environmental Issues by Education Levels

Dependent Variable: The Environmental issue in China is Terrible	
High school	0.133***
	(0.0122)
Some college or above	0.176***
	(0.0145)
p-values	0.002
t-values	-3.17
The average value for below high school	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). 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 Some college or above=High school; t-value: t-values of test of Some college or above=High school. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

Figure C2: Event Study of the Number of College Students Over Time in Cities That Adopted the Expansion



Notes: We test for pre-trends and dynamics of the college expansion IV in an event study framework. Outcome is number of new college students, relative to the year of establishment.

Table C4: Concerns and Actions Taken on Environmental Issues

Panel A: Concerns on Environmental Issues

	Discuss environmental issues	Donation for environment protection	Concern over environmental issue
High school	0.134***	0.0585***	0.120***
	(0.0171)	(0.0115)	(0.0154)
Some college or above	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
High school	0.0255***	0.119***	0.0690***
	(0.00845)	(0.0140)	(0.0118)
Some college or above	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). 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 Some college or above=High school; t-value: t-values of test of Some college or above=High school. Standard errors clustered at the city level are reported in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$

D Appendix: Additional Counterfactual Tables

Table D1: The Productivity Effect of Reducing Pollution in Beijing (no externalities)

	Change in GDP per Worker (%)		
	Overall effects	Health	Relocation
Reduce steady state PM2.5	7.992	3.613	4.227
Relax skilled hukou	7.925	0.000	7.925
Relax unskilled hukou	-8.621	0.000	-8.621
Reduce PM2.5 & relax skilled hukou	15.388	3.613	11.364
Reduce PM2.5 & relax unskilled hukou	-0.545	3.613	-4.013

Notes: In this counterfactual exercise we reduce the steady state amount of pollution in Beijing by 50% (row 1). Next, we relax the *hukou* restrictions by skill level (rows 2 and 3) to the extent that the stringency of *hukou* regulation is as same as its median level of all cities. Finally (rows 4 and 5) we relax the hukou regulation to the same as its median level in China while reducing steady state pollution. Column 1 shows the gain to overall GDP per worker. Column 2 shows the component purely explained by the health-productivity channel. Column 3 through the pure relocation channel.

Table D2: The Productivity Effect of Relocating Pollution (no externalities)

	Change in GDP per Worker (%)		
	Overall changes	Health	Relocation
Relocate steady state PM2.5	5.437	1.761	2.826
Relax hukou	3.063	0.000	3.063
Relax overall mobility constraints	6.319	0.000	6.319
Relocate PM2.5 & relax hukou	8.261	1.761	5.545
Relocate PM2.5 & lower migration costs	11.834	1.761	8.865

Notes: In this counterfactual exercise we relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relocate only the exogenous component of pollution. In addition to such relocations of pollution, we also relax the *hukou* restriction in the 35 high tier cities to be the same level as the median city in the country (row 3). In row 4 we relax overall migration costs to the 35 high tier cities to be the same as the median city. Column 1 shows the overall gain to GDP. Column 2 shows the increase in GDP as a consequence of the health effects only. Column 3 shows the gain due to the re-allocation of labor channel only.

E Appendix: Additional Model Derivations

E.1 Deriving Labor Supply and Welfare

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

$$V_{jsod} = \epsilon_{jsd} w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_d exp^{-M_{sod}} \tag{7}$$

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

$$\begin{aligned}
\pi_{so1} &= Pr\left[\widetilde{w_{so1}}\epsilon_{s1} > \widetilde{w_{sod'}}\epsilon_{sd'}\right] \quad \forall d' \neq 1 \\
&= Pr\left[\epsilon_{sd'} < \frac{\widetilde{w_{s1}}\epsilon_{s1}}{\widetilde{w_{sod'}}}\right] \quad \forall d' \neq 1 \\
&= \int \frac{dF}{d\epsilon_{s1}} \left(\epsilon_{s1}, \omega_{so1}\epsilon_{s1}, ..., \omega_{soD}\epsilon_{sD}\right) d\epsilon_{s1}
\end{aligned} \tag{A.1}$$

where we define $\omega_{sod} \equiv \frac{\widetilde{w_{sp1}}}{\widetilde{w_{sod'}}}$. We assume that the abilities are distributed with the following Frechet distribution:

$$F(\epsilon_{s1}, ..., \epsilon_{sD}) = exp\left\{-\left[\sum_{d=1}^{D} \epsilon_{sd}^{-\eta_s}\right]\right\} \tag{A.2}$$

So the derivative of the CDF is given by:

$$\frac{dF}{d\epsilon_{s1}} = \eta_s \epsilon_{s1}^{-\eta_s - 1} exp\left\{-\left[\sum_{d=1}^{D} \epsilon_{sd}^{-\eta_s}\right]\right\} \tag{A.3}$$

This derivative evaluated at $(\epsilon_{s1}, \omega_{so1}\epsilon_{s1}, ..., \omega_{soD}\epsilon_{sD})$, allows us to determine the probability of choosing destination 1, given by π_{1os}:

$$\begin{aligned}
\pi_{so1} &= \int \eta_s \epsilon_{s1}^{-\eta_s - 1} exp\left\{-\left[\sum_{d=1}^{D} \left(\omega_{sod}\epsilon_{sd}\right)^{-\eta_s}\right]\right\} d\epsilon_{s1} \\
&= \frac{1}{\sum_{d=1}^{D}\omega_{sod}^{-\eta_s}} \int \left(\sum_{d=1}^{D}\omega_{sod}^{-\eta_s}\right) \epsilon_{s1}^{-\eta_s - 1} exp\left\{-\left[\epsilon_{s1}^{-\eta_s - 1}\left(\sum_{d=1}^{D}\omega_{sod}^{-\eta_s}\right)\right]\right\} d\epsilon_{s1} \\
&= \frac{1}{\sum_{d=1}^{D}\omega_{sod}^{-\eta_s}} \int dF(\epsilon) \\
&= \frac{1}{\sum_{d=1}^{D}\omega_{sod}^{-\eta_s}} .1 \\
&= \frac{\left(\widetilde{w_{so1}}\right)_s^{\eta}}{\sum_{d=1}^{D}\left(\widetilde{w_{sod}}\right)_s^{\eta}}
\end{aligned} \tag{A.4}$$

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 $\sum_{d=1}^{D}\omega_{sod}^{-\eta_s}$. Expanding on the definitions for $\widetilde{w_{sod}}$, and scaling up the probability by the size of the skilled workforce P_{os} by origin, we derive labor supply by skill and destination:

$$\pi_{sod} = \frac{\left[w_{sd} Z_d^{\gamma_s} hp_d^{-\nu_s} a_{sd} exp^{-M_{sod}}\right]^{\eta_s}}{\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{sd'} exp^{-M_{sod'}}\right)^{\eta_s}} \text{ and } L_{sd} = \sum_o P_{os}\pi_{sod} \tag{9-10}$$

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

$$\begin{aligned} E[V_{jsod}|d] &= (\widetilde{w_{sod}})\, E[\epsilon_{jsd}|d] \\ &= (\widetilde{w_{sod}})\, \pi_{sod}^{-\frac{1}{\eta_s}} \Gamma\left(1 - \frac{1}{\eta_s}\right) \\ &= \left(\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{sd'} exp^{-M_{sod'}}\right)^{\eta_s}\right)^{\frac{1}{\eta_s}} \Gamma\left(1 - \frac{1}{\eta_s}\right), \end{aligned} \tag{A.5}$$

where Γ is the gamma function, and is constant across cities.

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 those who originate from this city already have access to the amenities without paying *hukou* costs. We define average utility for those from city o to be:

$$\overline{V_{so}} \equiv \left(\sum_{d'} \left(w_{sd'} Z_{d'}^{\gamma_s} hp_{d'}^{-\nu_s} a_{sd'} exp^{-M_{sod'}}\right)^{\eta_s}\right)^{\frac{1}{\eta_s}} \tag{A.6}$$

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_s as it captures the dispersion in tastes across locations. The utility of those in city o 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, by using the above set of equations:

$$log\ \pi_{sod} = \eta_s log\overline{V_{so}} + \eta_s\left(log\ w_{sd} - \nu_s log\ hp_d\right) + \eta_s log\ a_{sd} + \eta_s\gamma_s log Z_d - \eta_s M_{sod} \tag{11}$$

E.2 Elasticity of Capital, and Modelling Skill-biased Capital

So far the model assumes that capital is perfectly supplied at the rate R^*. If however, capital was fixed at a value $\bar{K_d}$ in a city, it would not change the qualitative predictions of the model. The average earnings for a worker with skill s in district d would be:

$$\log w_{sd} = \log\left(\frac{\partial Y_d}{\partial \ell_{sd}}\right) = \log\theta_{sd} + \left(\left(\frac{1}{\sigma_E} - 1\right)\left(\frac{1}{\varrho}\right)\log Y_d - \left(\frac{1}{\sigma_E} - 1\right)\left(\frac{1-\varrho}{\varrho}\right)\log\bar{K_d}\right) - \frac{1}{\sigma_E}\log L_{sd} \tag{A.7}$$

Here the modified term $\left(\frac{1}{\sigma_E} - 1\right)\left(\frac{1-\varrho}{\varrho}\right)\log\bar{K_d}$ is common across skill levels and constant.

We can explicitly model skill biased capital as affecting the productivity parameter θ_{sd}. Below, we explicitly model skill biased capital to show how flexible forms of introducing it do not influence the estimation. In the following set up, the noticeable changes are where Equation

4 has been modified into Equation A.10, which includes an elasticity of substitution between labor ℓ_{sd} and skill biased capital k_{sd} represented by σ_s:

$$Y_d = A_d L_d^{\varrho} K_d^{(1-\varrho)} \tag{A.8}$$

$$L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E - 1}} \tag{A.9}$$

$$L_{sd} = \left(\Lambda_s k_{sd}^{\frac{\sigma_s - 1}{\sigma_s}} + (1 - \Lambda_s) \ell_{sd}^{\frac{\sigma_s - 1}{\sigma_s}} \right)^{\frac{\sigma_s}{\sigma_s - 1}} \quad , \tag{A.10}$$

where ℓ_{sd} is the supply of workers of skill s, and L_{sd} is now a labor aggregate over workers and capital. Given this new set up, earnings can be represented by Equation A.11, instead of Equation 5:

$$\log w_{sd} = \log \left(\frac{\partial Y_d}{\partial L_{sd}} \right) = \frac{1}{\varrho} \log A_d + \log \widetilde{\varrho} + \log \theta_{sd}(1 - \Lambda_s) + \frac{1}{\sigma_E} \log L_d + \left(\frac{1}{\sigma_s} - \frac{1}{\sigma_E} \right) \log L_{sd} - \frac{1}{\sigma_s} \log \ell_{sd} \ , \tag{A.11}$$