The Productivity Consequences of Pollution-Induced Migration in China

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

June, 2023

Abstract

We quantify how pollution affects aggregate productivity and welfare in spatial equilibrium. We document a robust pattern in which skilled workers in China emigrate away from polluted cities, more than the unskilled. These patterns are evident under various empirical specifications, such as when instrumenting for pollution using distant upwind power-plants, or thermal inversions that trap pollutants. Pollution changes the spatial distribution of skilled and unskilled workers, which increases returns-to-skill in cities that the educated emigrate from. We quantify the loss in aggregate productivity due to this re-sorting by estimating a spatial equilibrium model. Counterfactual simulations show that reducing pollution increases productivity through spatial re-sorting by approximately as much as the direct health benefits of clean air. We identify a new channel through which pollution lowers aggregate productivity significantly. *Hukou* mobility restrictions exacerbate welfare losses. Skilled workers' aversion to pollution explains a substantial portion of the wage-gap between cities.

Keywords: Spatial productivity gaps, pollution and migration, labor reallocation
JEL Codes: E24, 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
[§]National University of Singapore, ran.song@yale-nus.edu.sg

We thank seminar participants at UCSD, UW (Seattle), NBER Summer Institute (Energy and Enivornmental Economics; and Urban Economics), NBER BREAD Junior Conference, Cowles Foundation Conference, LSE Workshop in Environmental Economics, Stanford SITE Migration Conference, U Richmond, UC Berkeley, U of Virginia, NUS, Cities and Development (Johns Hopkins), China Conference on Growth and Development, Ortygia Business School, Notre Dame, Georgetown, Barcelona UAB, the World Bank, Harvard, CUHK, HKUST (China Summer Institute), Nankai-Yale Conference, Peking U (CCER Summer Institute; NSD), Xiamen U and Lingnan U for useful comments, and to Ben Faber, Ed Glaeser, Matthew Kahn, David Lagakos, Joan Monras, Melanie Morten, Paulina Oliva and David Yang for insightful suggestions.

1 Introduction

The large productivity gaps across regions within countries create an enduring development puzzle: Why do workers remain in low productivity areas when they could experience wage gains elsewhere (Gollin et al., 2014)? We focus on air pollution, which can have a large effect on where people choose to live and work. Particulate matter pollution exceeded WHO air quality guidelines for 96% of Chinese cities in 2015, and was on average four times higher than safe levels, adversely affecting life expectancy, heart disease, stroke, and lung cancer (Ebenstein et al., 2015, 2017; He et al., 2020). This recent sharp increases in pollution was concentrated in a few Chinese cities, altering incentives for workers to relocate between locations.

We quantify the aggregate productivity consequences of this movement within a spatial general equilibrium framework that allows for rich policy counterfactuals. Policies that permit spatial reallocation could potentially produce productivity gains (Clemens et al., 2019; Heise and Porzio, 2021). We show that the emigration of workers from polluted, but potentially highly-productive cities, creates losses in aggregate productivity. Additionally, when skilled workers choose to leave polluted places where they would be more productive, the production complementarities between skilled and unskilled workers makes the unskilled less productive. Migration costs, both physical and from Chinese *hukou* policy, differentially restrict mobility by skill, and exacerbates the productivity and welfare losses from pollution.

We identify the migration response to *exogenous variation* in pollution. We go beyond the existing reduced-form literature by assembling several new datasets and investigating this relationship under multiple independent sources of variation, to robustly establish that skilled people indeed emigrate away from polluted areas. We isolate exogenous fluctuations in pollution leveraging variation in wind direction combined with the historical placement of distant thermal power plants (as in Freeman et al., 2019), and a meteorological phenomenon called thermal inversions that traps pollution (as in Arceo et al., 2016; Chen et al., 2022; Hicks et al., 2015). We find robust evidence that workers leave areas with high levels of pollution, and college-educated workers are more responsive.

Yet, the reduced form analysis by itself is incomplete due to equilibrium effects: all parts of the country are affected either directly or indirectly by the relocation of workers, making finding true 'control groups' for comparison elusive. Pollution and the skill-composition of the workforce are jointly determined, and both depend on other factors such as industrial growth (Bassi et al., 2021). Residents of cities experiencing no change in pollution will still see changes in migration incentives, as immigration from other newly-polluted places change equilibrium wages across cities. As the quantities of workers, wages, and pollution levels are jointly determined in spatial equilibrium, it is difficult to interpret key parameters solely from reduced-form relationships. Our model interprets the estimated parameters and allows us to study policies.

We quantify the productivity implications using our model. The differential emigration of skilled workers changes wages by skill-group across cities, which in turn also affects the location choices of the unskilled in general equilibrium (Eeckhout et al., 2014). We empirically document in our data that the relative scarcity of the skilled 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 to high marginal product areas, which raises aggregate output (Hsieh and Moretti, 2018).

The differential response to (exogenous variation in) pollution shifts the labor supply of different types of workers, and produces a valid estimate of the compensating wage-differential for pollution. This traces out the labor demand curve, and allows us to estimate the elasticity of substitution between skills, showing a fair degree of production-complementarity. To trace out the labor supply curve, we create instruments based on trade-induced growth from the permanent normalization of trading relations (NTR) between the US and China which differentially generate variation in the demand for skilled and unskilled workers across Chinese cities.[1]

For a more comprehensive and accurate quantification, our model incorporates alternative mechanisms linking production, pollution and productivity: (a) pollution can directly affect health and productivity (He et al., 2016; Zivin and Neidell, 2012); (b) production, in turn, affects air quality; and (c) worker location decisions may affect agglomeration (Au and Henderson, 2006), (d) house prices (Glaeser, 2014a), or (e) the pollution-intensity of production (Glaeser and Kahn, 2010; He et al., 2020). We introduce additional instruments

[1] Pierce and Schott (2016) use the NTR shock and Autor et al. (2013) use the WID shock to document effects on the US. Our approach takes advantage of the fact that these are simultaneously export shocks that had differential effects on skilled and unskilled labor demand in Chinese cities that were more or less exposed to trade. 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.

to estimate these elasticities.[2] As workers move, it may change where production takes place (Imbert et al., 2022) and the sectoral composition of production (Duranton and Puga, 2005), affecting pollution levels, agglomeration and house prices. In summary, we quantify the productivity effects of pollution via worker re-sorting, *accounting for* other important mechanisms through which production, pollution, and migration are related.

The model allows us to quantify how much of the wage gap across cities is attributable to pollution differences. Our estimates imply that equalizing pollution between high-pollution Tianjin and low-pollution Chongqing would bridge the between-city skilled wage gap by 14%. Companies in China reportedly offer up to 20% wage premiums to induce workers to relocate to polluted productive cities (AFP News, 2019), so our estimates are in line with the real-world behavior of firms and workers.

The fact that pollution explains a meaningful portion of the productivity gaps across cities sheds light on the puzzle we raised at the outset. The news media has documented how concerns about pollution keep workers away from cities where they could be more productive,[3] but such productivity losses have not been quantified. This phenomenon is also not limited to China: when Delhi residents were asked about their plans to deal with pollution, the single-most common response was "relocate" (Kapur, 2019). Pollution also affected within-city neighborhood choice in 19th century Britain (Hanlon, 2020; Heblich et al., 2021), and recent reports of emigration following wildfires in California suggest that this may not be solely a developing world phenomenon.[4]

We use the model to evaluate the effect of counterfactual changes in environmental policy. In one exercise, we only halve the 'exogenous' component of pollution in Beijing (say, by relocating upwind coal-fired plants,or investing in greener technologies), but allow for pollution to rise as workers correspondingly migrate in. In another counterfactual we halve the amount of total steady-state pollution (say, by setting pollution caps for Beijing),

[2]Our estimated elasticities are similar to credible estimates in the literature on the direct effect of pollution on productivity (Adhvaryu et al., 2022; Chang et al., 2019; Kahn and Li, 2019), and worker location on agglomeration (Moretti, 2004). If we were to discipline our model by calibrating elasticities from the literature we would get similar quantitative results.

[3]See for instance, "*Why leave job in Beijing? To breathe*." Wall Street Journal, April 14 2013. https://www.wsj.com/articles/SB10001424127887324010704578418343148947824, and also "*Execs fleeing China because of bad air*," CBS news, Jan 29 2013, https://www.cbsnews.com/news/execs-fleeing-china-because-of-bad-air/.

[4]See "*How Climate Migration will Reshape America*," New York Times, https://www.nytimes.com/interactive/2020/09/15/magazine/climate-crisis-migration-america.html.

which would also be a function of endogenous feedback loops in our model. In each scenario, incomes rise by more than 12%, mostly as a consequence of skilled workers moving into Beijing.

We then investigate the effects of changing the spatial distribution of pollution within the country (but holding total country-wide emissions constant) by moving pollution away from cities with more skill-biased capital. Setting city-specific pollution caps increases GDP by 6.7%, while relocating coal-fired plants away from such cities raises aggregate Chinese GDP by 3.67%. Relaxing migration costs (less stringent *hukou*) amplifies effects.

Pollution-control programs recently introduced by the Chinese government suggests that authorities already recognize such gains. The *12th Five-Year Plan* sets targets for ambient concentrations of particulate matter, with more stringent targets for high-productivity, polluted regions like Beijing.[5] We use our model to quantify the consequences of this policy. Despite targeting only a subset of cities, this increases aggregate GDP in China by 3.6%. However, our simulations suggest the policy largely benefits skilled workers. Reducing migration barriers in conjunction with pollution control would result in a more equitable distribution of benefits across skilled and unskilled workers.

All these model simulations show that relocation of workers is a major driver of these effects, as large as the direct effect of air pollution on health and productivity. The relationship between pollution and health is the subject of a large literature in economics and epidemiology, but we learn that ignoring labor mobility grossly underestimates the overall consequences of pollution on an economy's prosperity. This is relevant for policy, as pollution and migration have been two of the defining features of Chinese growth (Brandt et al., 2008; Tombe and Zhu, 2019; Zheng and Kahn, 2013).

Whether relocating pollution affects aggregate welfare (beyond productivity) depends on the underlying reason as to why the high and low-skilled react differently. Survey data shows that this is partly due to different preferences and environmental awareness of the rich, and partly due to differential costs of migrating. Chinese cities have a point-based system that exempts workers with skills or higher education from their *hukou* restrictions (Table C1). Without the exemption, the system imposes a burden on poor in-migrants to cities by limiting or prohibiting their access to many government-provided benefits (Combes et al., 2019). With high mobility costs, unskilled workers may be trapped in low-wage pol-

[5] http://www.mep.gov.cn/gkml/hbb/bwj/201212/t20121205_243271.htm, accessed 9/17/19.

luted cities even as their skilled counterparts leave, and our modeling shows that welfare losses from pollution are magnified when migration costs are high.

Our work is related to several literatures and sub-disciplines in economics. We are among the first to document how pollution and mobility interact to lower aggregate output. While local development planners and news reports have worried that pollution causes 'brain drain,' our spatial general equilibrium analysis quantifies the productivity consequences of such migration, and speaks to the urban and macroeconomics literatures. Second, environmental economists have focused on how pollution lowers productivity via threats to health, but we show that at least as important is the productivity effects of worker re-sorting. This suggests that the productivity consequences of pollution may be twice as large as what environmental and health economists previously taught us. And third, we contribute to the long-standing debate among macroeconomists and development economists on what explains the productivity gaps across regions within countries. Pollution explains a substantial portion of these gaps in China.

Our primary contribution is in quantifying the aggregate productivity consequences and studying policy counterfactuals. Other contemporaneous papers documented a reduced form relationship between pollution and migration (Chen et al., 2022; Chen, 2023; Yu et al., 2020). For example, Chen et al. (2022) infer migration responses to air quality 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. While not our direct aim, we do contribute to the reduced-form literature in a few ways. First, we use the 2015 Population Census of China on *actual* migration decisions (after pollution information was widely disseminated), and longitudinal panels which track individual migration over time from 2008-16 (before and after pollution information was disseminated) to explore the relationship. Second, we use multiple sources of policy-relevant variation, and a battery of sensitivity tests to build confidence in the pollution-migration relationship. Third, we additionally document the changes in net populations similar to Chen et al. (2022), but recognize them to be the consequence of re-sorting across all cities in general equilibrium. Indeed, without a model, it is challenging to interpret reduced-form population changes as all cities are indirectly affected, and there are no pure 'control cities.' Different data and empirical strategies used by these papers produce results consistent with the mechanism highlighted in our paper.[6] But, our model helps us answer questions different from all other work: what are the macroeconomic consequences of pollution on production and economic well-being, and what are the consequences of pollution control?

While rich, in some respects our model is intentionally parsimonious. We aim to provide a tractable framework, and estimate each parameter from within our data, rather than calibrate it from other settings. Despite parsimony, in validation exercises, we match major empirical patterns, suggesting these abstractions are unlikely to have implications.

2 Data and Measurement

2.1 Demographic and Migration Data

We measure internal migration using the 2015 Population Census of China. This is the latest census with restricted public access, and importantly, the only population census after both the 2008 disclosure of PM2.5 data by the US Embassy in China and city-level PM2.5 data by the Chinese Government in 2012 (Appendix B). The census records demographic and economic characteristics of individuals, including education, employment, *hukou* location, and current residential city. We restrict our attention to the working age population.

We define migration in a few ways. First, in the Census, migrants are defined as those who are away from their *hukou* city for more than six months.[7] *Hukou* status determines citizens' access to state goods (e.g., schooling) and services (e.g., marriage registries or passports).[8] Given the strong (forced) attachment to one's *hukou* city, when a person's location of residence differs, it can be reliably characterized as migration.

Second, we construct an individual-level longitudinal panel using the China Labor-

[6] Xue et al. (2021) show that polluted Chinese cities experience drops in skilled executives once pollution data was made public. Other work documents Chinese households' willingness to pay to avoid pollution using variation in housing prices (Freeman et al., 2019) and air filters (Ito and Zhang, 2019).

[7] This definition is consistent with other recent work on internal-migration in China (Combes et al., 2019; Tombe and Zhu, 2019). Most people's *hukou* city is their birth city. In the 2014 CLDS only 7% of respondents's a *hukou* city was different from their birth city. Additionally, it usually takes a long time for migrants to obtain local a *hukou*. Some local governments require that migrants must work in the city for more than 3 years before applying for local a *hukou*.

[8] *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. 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.

force Dynamics Survey (CLDS), which records individual histories of location changes. The CLDS is a national longitudinal social survey, with information on education, work and migration experience. Since the survey asks retrospective migration histories of each individual, we are able to construct a longitudinal panel of locations between 2008 and 2016. We define migration to be an indicator for whether an individual changed cities, regardless of *hukou* status. The CLDS allows us to account for individual-specific unobservables, track those who have moved multiple times and those who have returned.

We supplement migration flows with the *net* stock of workers by skill level computed using the Census. These changes in stock are the summary outcome of (net) migration decisions for all reasons and through all modalities (whether or not individuals change *hukou* status), and the object most sensible to use in our structural analysis for the quantification of productivity. Our structural quantification helps account for the fact that changes in the net-stock of workers cannot simply be interpreted as a migration response to pollution. Jointly, the three different migration measures either follow best practice, or improve on the approaches in the existing literature to measure migration in China. Since the 2015 Census does not record individual-level wages, we use the CLDS to calculate city-and-education specific average wage.

2.2 Air Quality, Inputs into Instrumental Variables and Controls

We use satellite data to measure air quality, which has a few advantages over official sources. First, satellite-based PM2.5 measures are available for all cities between 1998 and 2015, whereas official PM2.5 data are only available since 2012.[9] Second, official air quality data may be subject to manipulation by local governments (Ghanem and Zhang, 2014; Greenstone et al., 2020). City-level annual PM2.5 concentrations are measured using the Global Annual PM2.5 Grids derived from satellite data (Van Donkelaar et al., 2016).

We obtain information on large-scale (capacity>1 million KW) power plants and their coal consumption from China's Electric Power and Energy Statistical Yearbooks. We supplement this with information on the establishment year of plants, the angle between their locations and annual prevailing wind direction of each city, and distance to each city.

We collect data on thermal inversions from the Modern-Era Retrospective Analysis for

[9]PM2.5 is hazardous, considered to be the best indicator for pollution health risks, and has become the dominant air pollutant in China in recent years (Barwick et al., 2019). We use the Air Quality Index (AQI) released by Ministry of Environmental Protection (MEP) for robustness.

Research and Applications (MERRA-2). 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.

Estimating the structural model requires us to develop a few other instruments. First, 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. Second, we leverage variation in trade shocks to identify migration responses to wages. Trade data are from the China Customs Database.

We collect city characteristics, such as population and GDP, from the City Statistical Yearbooks. Weather data come from the Meteorological Data Service Center. We provide more details on all data sets in a Data Appendix F.

3 Empirical Strategy

Our data allow us to examine pairwise flows between cities, as we do when we estimate our labor supply curves in our general equilibrium analysis. Our regression analysis will begin with the simplest specification: the effects of PM2.5 concentration in origin city o on the amount of out-migration by skill group.

$$M_{io} = \alpha + \beta_1 Log(PM2.5)_o + \mathbf{X}\beta + \varepsilon_{io} \ , \tag{1}$$

where M_{io} is an indicator for whether or not individual i left origin city o, and $\mathbf{X}$ are controls. The dependent variable is the actual migration decision (i.e., flows, rather than stocks). After presenting the model, we show results using the net stock of workers by skill. Changes to the net stock are difficult to interpret in reduced form without the aid of the model, as the quantities of workers are determined in spatial general equilibrium.[10]

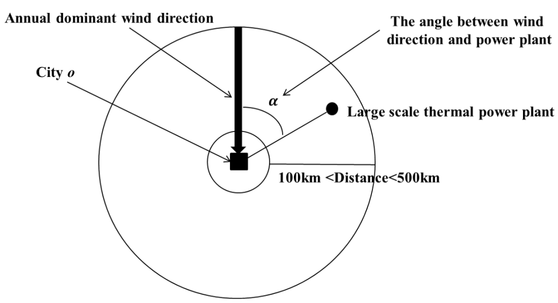
We use a few different identification strategies to isolate the causal effect of pollution on migration: Panel fixed-effects models, as well as two different instrumental variables strategies. We examine the same relationship using multiple data sources, different types

[10]When using net quantities of workers, there is no 'control group,' as all movement out of a polluted city implies movements into a non-polluted city, violating the Stable Unit Treatment Value Assumption (SUTVA). Our GE model helps interpret magnitudes in such a setting. The model also derives analogous bilateral origin-destination and destination-level regressions, which we estimate after describing the underlying theory.

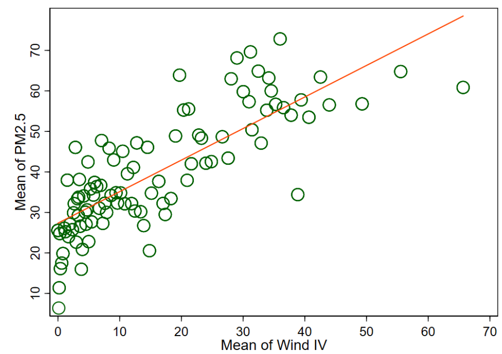
of variation, and conduct specification tests, falsification tests, sample restrictions, and different types of controls (see Appendix A and B).

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

Figure 1: 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 o. The dark dot represents a large-scale thermal power plant located at least 100km outside city o and within 500km from the city. The angle α denotes the angle between the annual prevailing wind direction of city o and the large-scale power plant. Large-scale thermal power plants are defined as 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 in 2015 and x-axis denotes the mean value of wind direction IV in each quantile.

Our first source of variation measures the extent to which each city is down-wind of distant coal-fired power plants (Freeman et al., 2019). The instrument is a function of wind direction and coal consumption in large-scale thermal power plants located in a 100-500km radius around the city, penalizing plants farther away, and those located upwind of the city:

$$Log(PM2.5)_o = \gamma_0 + \gamma_1 \sum_{p}^{P} \left(\frac{1}{\alpha_{po}+1} \right) \left(\frac{1}{dist_{po}} \right) C_p + \varepsilon_o \, , \tag{2}$$

where α_{po} denotes the angle between the annual prevailing wind direction of city o and the plant p, $dist_{po}$ is the distance between the plant p and city o, C_p is the annual coal consumption in plant p. Figure 1a explains the intuition. The underlying variation is driven by how wind patterns blow pollutants from distant coal plants to cities. Indeed, as we later show, the primary driving factor is simply the wind direction. Our first-stage relationship in

Figure 1b shows that cities downwind from, and closer to, higher coal-consumption power plants are more polluted.

We expect our instrument to be orthogonal to local economic activity. Large-scale thermal power plants supply electricity to vast areas of China; many do not supply to nearby cities, but rather to remote provinces (Freeman et al., 2019). Inter-province electricity transmission is prevalent in China. For example, in 2016, Yunnan province transmitted about half of its power output to other provinces, and Sichuan province transmitted 40% of its power output to other provinces. Large-scale power plants are mainly responsible for the power generation for inter-province electricity transmission in China (Freeman et al., 2019). Local governments find it difficult to exert influence the setup of large plants, their siting decisions, or allocation of electricity from them. The impact of distant plants on local economic activity is thus small, and we chose to focus on plants at least 100km away to be cautious.

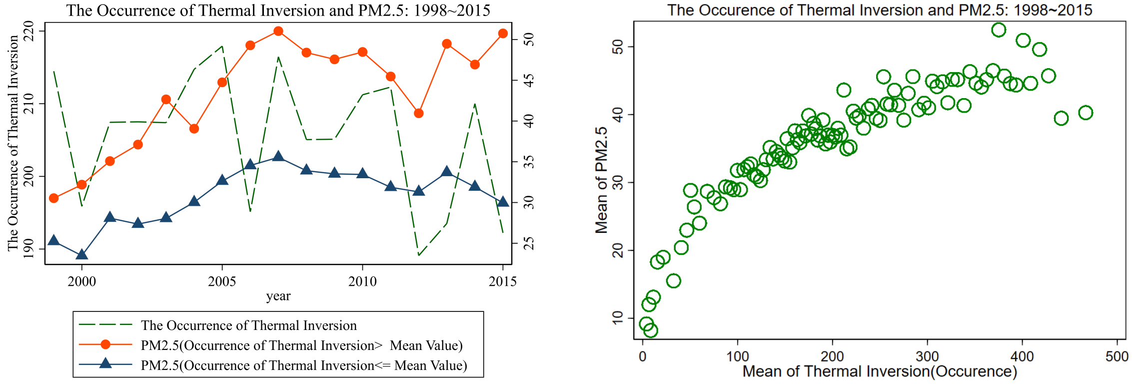
3.2 Instrumental Variable 2: Thermal Inversions

Our second instrument uses the number and strength of thermal inversions, a meteorological phenomenon where the above-ground temperature is abnormally higher than the ground temperature, trapping pollutants. This has been used as an instrument for air quality in Mexico (Arceo et al., 2016; Molina, 2021), the US (Hicks et al., 2015) and Sweden (Jans et al., 2018), among other places. Most recently, Chen et al. (2022) show that the number of thermal inversions predicts the movement of people across China. We use newer migration data from the 2015 Census at the individual level, rather than quantifying migration indirectly from population changes.

We create two measures of inversions in city o: the number of thermal inversions in each year, and the annual mean strength of these inversions.[11] As polluting potential rose over time in China, areas with more thermal inversions trapped pollutants in the nearby atmosphere (Figure 2). The right panel shows a strong first stage. In our specifications we control for time-varying weather amenities, show variants of our measures of inversions, and show how past pollution does not predict future inversions.

[11] The literature argues that inversions are assumed to be exogenous for three reasons. First, temperatures from one year to the next within a region are considered to be exogenous (Dell et al., 2012). Second, in the atmospheric dispersion literature, pollution is a function of winds, settling and emissions; whereby inversions reduce atmospheric ventilation and traps pollutants (Sharan and Gopalakrishnan, 2003). Third, inversions are a function of chemical potential and natural parameters (Ferrini, 1979) .

Figure 2: Thermal Inversions and Air Quality



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 and x-axis denotes the mean occurrence of thermal inversions in each quantile.

4 The Relationship Between Pollution and Migration

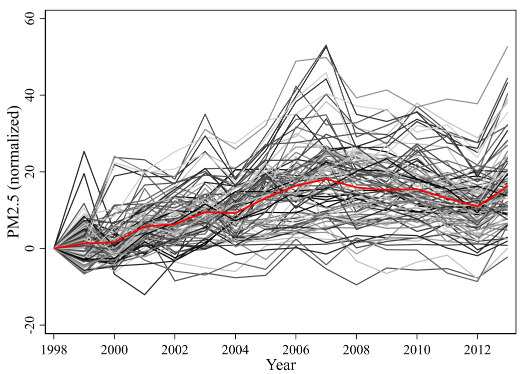
We first describe the spatial and temporal patterns of pollution, migration and wages in the raw data. Figure 3a plots the time trend of annual PM2.5 concentration by city (which exceeds WHO guidelines, every year). The figure also shows that the increase in the mean coincided with increases in cross-city dispersion of pollution (Figure C1a). The increase in the mean and variance was driven by dramatic increases in PM2.5 in a subset of cities. Figure C1b shows that both wages and the cross-city variance in wages rise over time. If this implies an increase in the spatial dispersion of marginal products of labor, then it raises the possibility that moving workers from low marginal product to high marginal product cities increases aggregate output.

Figure 3b shows the spatial variation in annual average satellite PM2.5 concentration for 2015. The north-east and east experience severe air pollution. The north-east further suffers from coal-burning for heating, whereas the east has manufacturing industries.[12] Correspondingly, low-skill emigration rates are high in the south of China (Figure 4a),

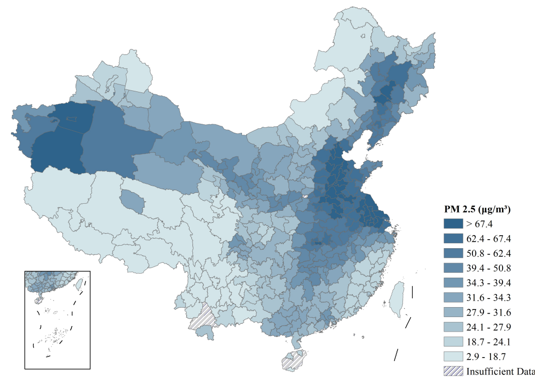
[12] Dust-storms in southern Xinjiang province are responsible for the isolated area of high particulate matter observed in the west. This area is otherwise not highly economically active.

Figure 3: The Distribution in Pollution Across Cities and Over Time

(a) Change in mean PM2.5 over time



(b) PM2.5 Concentration in China

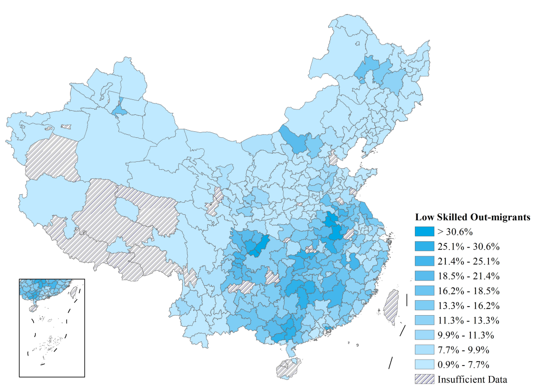


Notes: Spatial and temporal distribution of PM2.5 using the Global Annual PM2.5 Grids. The map shows the geographic spread in 2015. Figure 3a shows the increase in PM2.5 over time for the 100 largest cities, relative to the 1998 PM2.5 value (the difference with respect to 1998). The red line is the unweighted average.

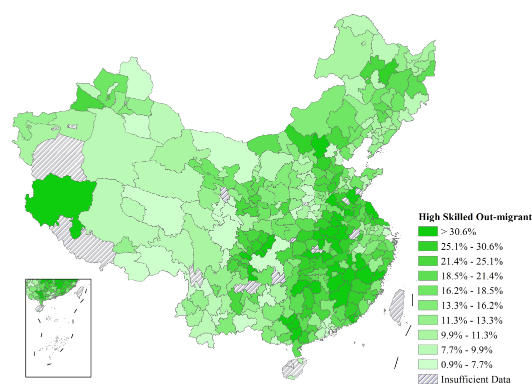
while high-skill out-migrants are comparatively more populous in the north-east and the east (Figure 4b). These figures therefore jointly suggest that pollution is more spatially correlated with high-skilled emigration rather than low-skilled.

Figure 4: The Geographic Distribution of the Share of Out-Migrants by Skill

(a) Share of Low-Skill Out-Migrants



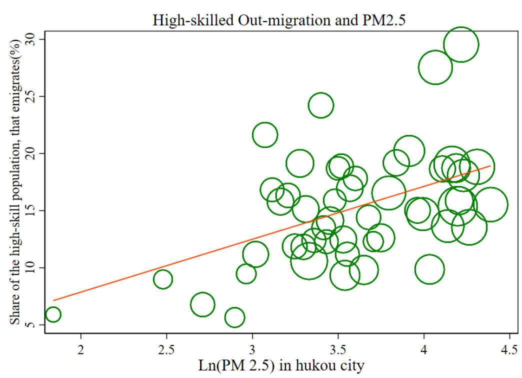
(b) Share of High-Skill Out-Migrants



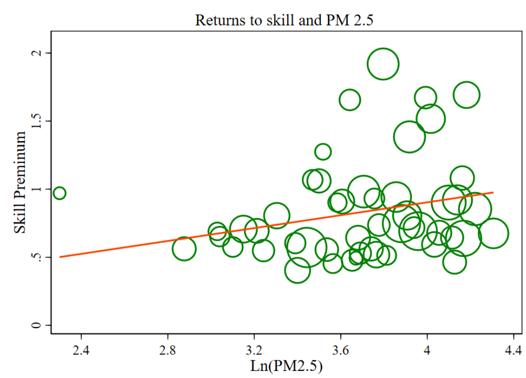
Notes: Low-skilled denotes people whose highest degree is high school or below. High-skilled have highest degree some college or above. Out-migrant shares are ratio of those who leave their *hukou* city for more than six months, and the number of people whose *hukou* location is a given city.

Figure 5a shows a clear positive association between pollution and high-skill emigration: the college-educated are more likely to leave polluted areas. Finally, Figure 5b shows that the wage returns to skills are higher in polluted cities, where skilled workers have emigrated from.[13] Figure 5b also highlights a key insight about the benefits of pollution control policy that will emerge in our model: Reducing PM2.5 in high-polluted cities induce high-skilled workers to move to where their skills are relatively scarce (and so their marginal product may be higher), and this re-sorting raises aggregate productivity.[14]

Figure 5: The Effects of High PM2.5 at Cities



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



(b) Returns to Skill and PM2.5

Notes: The share of high-skilled out-migrants is the share of some college (or above)-educated emigrants from the city-level college-educated *hukou* population. 'Returns to skill' denotes the return to some college or above education. Each bubble is a city. The bubble size is weighted by the population in 2000.

4.1 The Causal Effect of Pollution on Migration

In Table 1 we examine the effect of pollution on an individual-level binary indicator of leaving one's *hukou* city. We show results separately for those with some college degree or above, and those without. We control for demographics, and for distances to three large seaports to account for the spatial distribution of economic development.

The OLS estimates in Panel A suggest that air pollution leads to the emigration of all types of workers, and this effect is statistically stronger for those with higher education. Columns 4-6 employ our first instrument based on wind direction and distant coal-fired

[13]Our estimates of returns are consistent with recent estimates from other work (Giles et al., 2019).

[14]Our results confirm Au and Henderson (2006) who show that Chinese cities are 'undersized,' and so less urban immigration is not welfare-enhancing.

Table 1: Pollution and Out-Migration

Panel A	Dependent variable: Leave hukou city indicator					
	OLS Regression			Wind+coal IV		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)
Log(PM2.5)	0.0428***	0.0399***	0.0501***	0.0772**	0.0609	0.140***
	(0.00974)	(0.00993)	(0.0112)	(0.0389)	(0.0423)	(0.0382)
Coeff diff pval		0.00			0.00	
Observations	761,548	643,124	118,424	761,548	643,124	118,424
R-squared	0.029	0.030	0.045			
F-test of IVs				52.53	46.52	41.54

Panel B	Dependent variable: Leave hukou city indicator					
	Number of inversions			Strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)
Log(PM2.5)	0.0906***	0.0871***	0.112***	0.0779***	0.0740***	0.107***
	(0.0202)	(0.0211)	(0.0233)	(0.0228)	(0.0238)	(0.0279)
Coeff diff pval		0.00			0.00	
Observations	761,548	643,124	118,424	761,548	643,124	118,424
F-test of IVs	102.5	97.33	84.83	51.09	49.60	40.54

Notes: Individual level regressions in 2015 across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. 'Coeff diff pval' reports the p-value of a test of coefficient equality between the low and high educated groups, using the Fisher's permutation test (following Cleary (1999), Brown et al. (2010) and Keys et al. (2010)). This bootstraps to calculate empirical p-values that estimate the likelihood of obtaining the observed differences in coefficient estimates if the true coefficients are, in fact, equal. The 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 use thermal inversions. All panels have: (a) City controls which include the log distance to Shanghai seaport, to Tianjin seaport and to Shenzhen seaport, and (b) demographics which include age, age-squared, gender, marital status, and an urban hukou indicator. The bottom panel using thermal inversions also have (c) weather controls which include temperature, wind speed, sunshine duration and humidity.Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic.

power plants (Freeman et al., 2019). The differential impact by skill level increases in magnitude when using our instruments. A 10% increase in PM2.5 raises out-migration rates by 0.77 percentage points overall, with the effect being meaningfully larger for those with higher education (1.4 percentage points) than those without (0.61 percentage points). The high and low-skill emigration responses are statistically different.

In Panel B, we study how variation derived from thermal inversions affects out-migration. We follow the literature, and control for weather amenities (Arceo et al., 2016; Molina, 2021). Again, the emigration response is more pronounced for high-skilled workers. A 10% increase in PM2.5 leads to a 1.12 percentage point increase in emigration rates for those with higher education, but only 0.87 for those without.

Appendix A.1.1 provides further details on our IV strategies. Table A1 shows (strong) first-stages. Columns 1-3 of Table A2 estimate a stacked regression, interacting pollution levels with skill levels (imposing common coefficients for all controls), showing again that the effect on the high skilled is larger. Our baseline specification that estimates these responses by splitting the sample, implicitly includes skill-specific controls, which allow us to capture the fact that the importance of distance to large cities and demographics may differ by skill level. In columns 4-6 of Table A2, we disaggregate education levels into more fine-grained categories and see an education gradient in emigration. Table A3 combines multiple instruments, and conducts overidentification tests.

Table A4 shows that workers also pay attention to the pollution levels at destinations, in deciding where to migrate to. In other words, workers leave polluted areas, and also seek out less polluted cities. Again, like the emigration response, the immigration location choices are more sensitive to pollution among higher educated workers.[15] In Section 6.2, we estimate a more comprehensive specification, documenting the pairwise flows between cities instrumenting for pollution. The model allows us to also account for changes in pollution in response to migration, as migration itself may affect pollution. These facts discipline our model: Fewer skilled workers in a polluted city raise the skill wage, and so cleaning up polluted cities will move skilled workers to cities with a higher marginal product, increasing aggregate output via relocation

4.2 Alternative Specifications, Robustness and Heterogeneity

In Appendix A.1.2 we examine various threats to identification. We summarize the results of these exercises in Figure 6, where a simple pattern emerges: both the high and low skilled emigrate from polluted cities, but the effect on the high skilled is stronger. These relationships are stable across specifications.

First, we evaluate the claim that plants may be systematically built near poorer, less

[15]Together, the in-migration and out-migration affect net population changes in a manner only slightly smaller than document by (Chen et al., 2022).

influential cities, and so the instrument may be correlated with unobservable characteristics of nearby cities. In Table A5, we exclude nearby plants, either in a 200km or a 400km radius of a city, and find that, if anything, our results are more precisely estimated. We may also be concerned that new plants are subject to more regulation as the Chinese government only recently paid attention to environmental issues. So, in Table A6, we restrict our sample to only old plants, and find similar effects. Table A8 demonstrates that the variation of our wind direction and coal-plants IV is primarily driven by wind direction, rather than distance and coal consumption. Table A9 shows policy makers do not intentionally avoid locating power plants upwind of politically important or populous cities. We show that baseline population, GDP and electricity consumption do not predict future upwind plants, or future iterations of our IV (Table A10). Our results suggest that it is not that policymakers avoid richer, influential cities when building plants, and that plants are not built in areas with higher need for electricity, perhaps as most electricity is directly supplied to the larger grid.

In Table A11, we create 'placebo' instruments, artificially changing the wind direction and showing that these falsified instruments are less likely to predict pollution levels and migration decisions. Similarly, for our thermal inversion instruments, in Table A12 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 exploit a different source of pollution variation, driven by the Huai river heating policy (Chen et al., 2013) which generated an artificial discontinuity in air quality on two sides of the river. North of the river, the government encouraged centralized heating systems which primarily relied on coal-fired boilers, leading to a discontinuity in air quality across the river. While there is no differential out-migration in areas immediately north versus south of the river, skilled workers are discontinuously less likely to migrate into the more polluted cities just north of the river. We limit this discussion to the appendix, as the Huai river RD estimates a LATE difficult to relate to our model, as it may capture migration over short distances across the river.

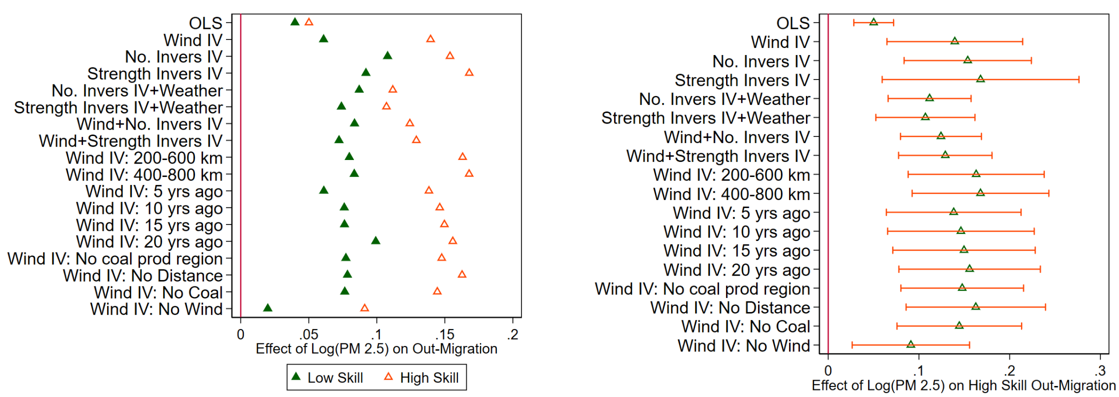
In Appendix A.2, we turn our attention to studying different model specifications, subsamples, and checking the robustness of our estimates to different controls. Figure 7 summarizes the results of this large set of robustness exercises.

First, in Appendix A.2.1 Table A13, we show similar patterns using an individual-level longitudinal panel data and a different definition of migration. The longitudinal panel allows us to track individuals' spatial sorting over time, and control for individual fixed effects. Importantly, we define migration to be an indicator for whether or not an individual changed their city, regardless of whether they change their *hukou*.

In Appendix A.2.2 Tables A14 we study cumulative pollution exposure and find that pollution exposure spread over a longer time period has a larger impact than shorter time frames. In Appendix A.2.3 Table A15 we firstly 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. We then drop coal producing regions as coal plants may locate near coal production. We find that youth are more responsive to pollution when making location choices (Table A16).

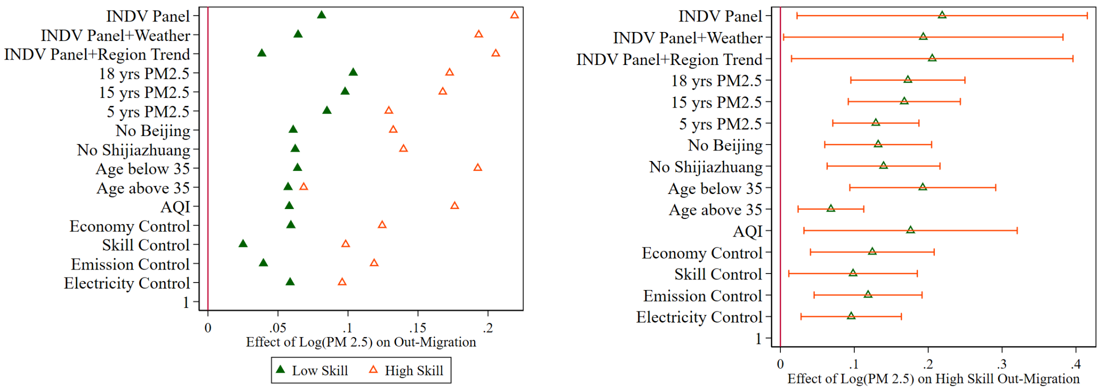
Table A17 uses the widely reported/publicized Air Quality Index (AQI) as the endogenous variable of interest, capturing the combined impact of many pollutants. In Appendix A.2.4 Table A18, we also control for annual rainfall, elevation, area under water, and coastlines. Table A20 highlights robustness to a long list of socioeconomic controls, including

Figure 6: Different Sources of Variation



Notes: Summary of Appendix A.1 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-skill workers, and include 95% confidence intervals. Instruments include number ("No. Invers" for short) and the strength of thermal inversions ("Strength Invers") as well as different versions of the wind direction and coal-fired power plants IV ("Wind IV"). "+Weather" include weather controls. "No Wind" excludes wind direction from the IV. "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 the city.

Figure 7: Different Models, Samples and Controls



Notes: Summary of results using different models, samples and controls. We compile coefficients from different specifications. On the left we show the coefficients on high and low skill workers. On the right, we focus on high-skilled workers, and include 95% confidence intervals. Individual longitudinal panel accounts for individual fixed effects ("INDV panel"), for local weather conditions ("INDV panel+weather") or region-specific trends ("INDV panel+region trend"). "5 yrs - 18yrs" aggregates past pollution into different time periods. "No Beijing" excludes Beijing. "AQI" is the Air Quality Index outcome. Different "Economy", "Skill", "Emission" and "Electricity" controls are described in Appendix A.2.4.

the skill-distribution, baseline economic indicators, and industrial pollutants.

Finally, in Appendix B, we show that the disclosure of official pollution data affected migration (Table B1). The striking consistency of the migration response to pollution, especially among the high-skilled, across a large set of robustness checks, many different, independent sources of variation in the data, different estimation strategies, samples and variable definitions, gives us confidence that we are indeed identifying an effect that is real.

4.3 Wage Returns and Pollution

This spatial re-allocation of skilled workers could produce a systematic relationship between pollution and the returns to skill across cities. If differential migration patterns alter the equilibrium stock of workers, then the relative scarcity of skilled workers in polluted cities may be associated with higher returns to skill. 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 have higher skilled and lower unskilled wages, and therefore higher skill returns.

Table 2 documents a simple empirical fact: As in the raw-data plot of Figure 5b, returns to skill are higher in polluted cities. Since, larger cities may be both more polluted and

Table 2: Pollution and Returns to Skill

Dependent variable: City-specific returns to college education				
	OLS		Wind+Coal IV	
	(1)	(2)	(3)	(4)
Log (PM2.5)	0.234*	0.428***	1.067***	1.974***
	(0.123)	(0.132)	(0.304)	(0.540)
Observations	129	129	129	129
R-squared	0.026	0.089		
City controls	No	Yes	No	Yes
F-test of IVs			39.41	22.29

Notes: City-level regressions of 130 cities weighted by 2000 population, for the sample of CLDS cities with non-missing skill-specific wage data. Robust standard errors are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal capacity of plant. City controls include the log distance to Shanghai, Tianjin and Shenzhen seaports. Returns to college education is calculated as the coefficient on 'some college or above' indicator in Mincer wage regressions controlling for age and gender. These regressions were done for each of the 130 cities in the CLDS data separately, as in Dahl (2002).

have higher premiums (Dingel et al., 2019), we use our instruments to address endogeneity concerns. We estimate the city-specific Mincerian returns using the CLDS data (as in Dahl (2002)), and explore the relationship between pollution and skill returns.[16]

4.4 Why are the High-skilled More Sensitive to Pollution?

Other than typical pecuniary migration costs, households must deal with institutional *hukou* policy which makes it difficult for unskilled workers to be mobile. In Table C1 we show how educational attainment can help gain *hukou* at certain destinations. Migration costs (both physical and institutional) may inihibit workers from accessing high-return locations, exacerbating productivity losses.

Different preferences or attachments may also drive the differential mobility responses we observe. Importantly, pollution concerns may loom larger for skilled workers, as the unskilled are focused on making ends meet with lower wages. Table C2 shows that educated workers are significantly more likely to discuss environmental issues, make donations for

[16]Even though the emigration of high-skill workers will tend to lower the low-skill wage, we should note that it is not necessary that the low-skill wage reduces with pollution. This is because the emigration of low-skill workers will tend to raise the low-skill wage, making the overall effect on low-skill wages ambiguous.

environmental protection, and appeal or raise concerns on environmental problems. The differences in migration patterns by skill therefore partly reflect differences in preferences, and partly differences in mobility restrictions.

5 Theoretical Framework

Our simple theoretical framework aids quantification of the productivity consequences of pollution-induced migration. Our model captures a few key features necessary for quantification. First, we endogenize the compensating differential, allowing pollution to have differential effects on the utility of skilled and unskilled workers. Second, mobility costs vary by skill. Together, these contribute to empirical patterns that show a differential out-migration by skill. Third, as college educated workers leave polluted cities, the marginal product of skilled labor rises. If skilled and unskilled workers are substitutes, the marginal product of unskilled work should fall. If there is some complementarity in production, the unskilled wage and skill-premium should rise (consistent with our observation that returns to skill are higher in more polluted regions).

We allow for additional channels as well. The changing structure of skills in a city affect production and pollution levels as population increases or the presence of more skilled workers may induce either more or less pollution-intensive industries to expand. This feedback effect of migration patterns on pollution emissions affects subsequent migration, which in turn affects production, and so on. Additionally, agglomeration may increase aggregate productivity if skilled workers converge, but house prices may also respond to such movements creating congestion. Finally, we allow pollution to affect the health (and lower the productivity) of all workers.

Our framework accounts for these feedbacks and generates estimable equations that we identify using instrumental variables. We summarize the model in a flowchart in Figure E1.

5.1 Production and Labor Demand

Aggregate output Y_d in destination city d depends on L_d (effective labor), K_d (capital), and A_d (TFP). TFP varies across cities, and depends on air quality Z_d and agglomeration forces. L_d depends on labor L_{sd} at each level $s = \{h, u\}$, high-skilled h and unskilled u.

$$Y_d = A_d L_d^{\rho} K_d^{(1-\rho)} \quad where \quad L_d = \left(\sum_s \theta_{sd} L_{sd}^{\frac{\sigma_E - 1}{\sigma_E}} \right)^{\frac{\sigma_E}{\sigma_E - 1}} \tag{3}$$

$0 < \rho < 1$ 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. We do not impose complementarity, and instead will be estimating σ_E.

Capital, K_d, is perfectly elastically supplied across cities at rental rate R^*, an assumption that can be relaxed.[17] We also allow for TFP A_d, to have a fixed city-specific component, essentially allowing a certain fraction of capital to be immobile (Brandt et al., 2013). The skill-biased productivity parameter θ_{sd} captures the productivity of each skill level. For instance, θ_{hd} increases with an increase in high-skill capital.[18] θ_{hd} also captures policies that raise wages for skilled workers. The value of θ_{sd} therefore varies across cities because of the variation in skill-biased capital, and other factors that make workers of a skill-group more productive in d. The average log earnings for skill s in destination d are:[19]

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

where $\log\widetilde{\rho} \equiv \left(\frac{1-\rho}{\rho}\right)\log\left(\frac{1-\rho}{R^*}\right)$ is common across all cities and workers. For tractability, output is a freely-traded numeraire, but housing is not traded across cities, and will have price effects across cities.[20]

Let us consider the determinants of the high-skilled wage w_{hd}. First, A_d is the amount of TFP in the city, which raises average earnings of all. Second, θ_{hd} 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_{hd}. As with any downward sloping demand curve, the more skilled workers there are, the lower the

[17]We show derivations and model extensions in our Appendix E.2.

[18]Skill-biased capital for high-skilled workers captures the presence of industries that hire college workers, such as finance, technology, skilled manufacturing, and professional services. For completeness, in our Appendix E.2, we explicitly model skill-biased capital within the nested CES framework and show how incorporating it does not affect the qualitative predictions. As we already model a freely traded K_d, we allow k_{sd} to be not tradable across cities, so that cities have certain inherent 'infrastructure' as well.

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

[20]We keep our focus on migratory frictions, and abstract away from location-specific trade costs (Tombe and Zhu, 2019), given the absence of high-quality cross-city trade data. Since we have only one traded good, trade costs are unlikely to affect our equilibrium. With multiple goods and trade frictions, more productive cities would have lower prices, and if anything, the welfare effects of moving pollution away from productive cities (and inducing in-migration) could be even larger.

skilled wage. Yet, L_d captures production complementarities, whereby an increase in the number of unskilled workers may raise the skilled wage. Equation 4 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. Migration of skilled or unskilled workers will change these quantities and affect skill-premia.

As workers move, this affects pollution levels, agglomeration and house prices, partially through changes in the sectoral composition of production (Duranton and Puga, 2005) and partially through different consumption preferences. First, in equation 5, each city has some 'exogenous' component of pollution $\bar{Z}_d$, that (say) depends on the occurrence of thermal inversions. Yet, equilibrium pollution levels Z_d depend on other endogenous workforce composition and production changes, which in turn raises pollution. The increase in pollution depends on both aggregate population (the size of the economy and industry, and congestion) (Rendon et al., 2014), and the skill mix, which change the type of production (industry vs services), amenities, energy consumption, or even local pollution policies:

$$Z_d = \bar{Z}_d \left(\frac{L_{hd}}{L_{ud}}\right)^{\psi_1} (L_{hd} + L_{ud})^{\psi_2} \tag{5}$$

We also allow for agglomeration economies, and for pollution to directly affect productivity. In equation 6, $\bar{A_d}$ is 'exogenous' city-level productivity (fertile soil, rivers, land, immobile capital, etc.), that does not move (i.e. a fixed factor of production). If $\phi_1 < 0$, pollution lowers the productivity of all workers, both high-skilled (Chang et al., 2019) and low skilled (Zivin and Neidell, 2012). We expect the number of skilled workers to raise TFP levels in the city via non-excludable innovation, such that $\phi_2 \geq 0$:

$$A_d^{\frac{1}{\rho}} = \bar{A_d} Z_d^{\phi_1} L_{hd}^{\phi_2} \tag{6}$$

5.2 Migration and Labor Supply

Workers have preferences over locations (be closer to home or certain cities). Indirect utility of worker j, with skill s, in destination d, from origin o is:

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

where μ_{jsd} is a random variable measuring preferences for a specific city d by individual j. A larger μ_{jsd} means worker j is particularly attached to city d. M_{sod} captures migration costs between o and d, including *hukou*, and physical costs that increase with distance. Migration costs vary by education level, and $M_{soo} = 0$ for natives. hp_d are housing prices, and v_s are the share of expenditures on housing by skill. [21] a_{sd} represents other non-pollution related skill-specific amenities of city d. ξ_{sod} (an error term) are differences in skill-specific amenities that depend on the origin. The compensating differential elasticity by skill is captured by γ_s.[22] Marginal workers are induced into migration by pollution, while inframarginal workers have higher utility in the city they live in.

We assume that μ_{jsd} are independently distributed and drawn from a multivariate extreme value distribution. The joint distribution of μ_{jsd} is given by:

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

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 strong, then workers are less likely to migrate in response to pollution.

Workers move to where their utility is higher. Given moving costs, there are no further arbitrage opportunities. Local ties and migration costs (including *hukou* and distance) are captured by μ_{jsd} and M_{sod} respectively. The share of workers with skill s from city o that move to d is:

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

Worker supply of skill s in city d–L_{sd} depends on the *hukou* origin population o–P_{os}. Thus, $L_{sd} = \sum_o P_{os}\pi_{sod}$. Taking logs of equation 9, we derive the labor supply curve:

$$log\,\pi_{sod} = -\eta_s log\overline{V_{so}} + \eta_s\left(log\,w_{sd} - v_s log\,hp_d\right) + \eta_s log\,a_{sd} - \eta_s\gamma_s logZ_d - \eta_s M_{sod} + \widetilde{\xi_{sod}}, \tag{10}$$

[21] Diamond (2016) and Piyapromdee (2021) also use a Cobb-Douglas utility function to analyze the migration choices of people and allow the share of housing expenditure in income to differ by skill in the utility function

[22] If $\gamma_h > \gamma_u$, good air quality is a normal good.

The error $\widetilde{\xi_{sod}}$ is $\eta_s log \xi_{sod}$. $\overline{V_{so}}$ captures the average utility of being from city o:

$$\overline{V_{so}} = \left(\sum_{d'} \left(w_{sd'} Z_{d'}^{-\gamma_s} hp_{d'}^{-\nu_s} a_{sd'} exp^{-M_{sod'}} \right)^{\eta_s} \right)^{\frac{1}{\eta_s}} \tag{11}$$

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, lowering wages and raising house prices. Yet, higher *hukou* restrictions will lower the utility for all individuals in other cities, as their option value of moving to a potentially desirable location falls. Therefore, $\overline{V_{so}}$ depends on city-specific mobility costs.

From equation 10 we see that η_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. Earnings differentials reflect compensating differentials (for pollution and other amenities) and migration costs.

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

5.3 Equilibrium and Elasticities

Equations 3-10 characterize 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. The model is characterized by a set of exogenous factors: city level producitivities $\bar{A_d}$, populations of the skilled and unskilled $\bar{L_h}$ and $\bar{L_u}$, 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 determine the quantities $A_d, Y_d, L_{hd}, L_{ud}, Z_d, K_d$ and prices $w_{hd}, w_{ud}, hp_d, R*$.

In equilibrium, the labor market clears for each skill level $\{h, u\}$. The supply of L_{sd} equals the demand for L_{sd} for all d, and all skills $\{h, u\}$. The sum of shares of migrant and non-migrants adds to one, or $\sum_d \pi_{hod} = \sum_d \pi_{uod} = 1 \; \forall \; o$. Output produced in a city is consumed in the city d, and there are no savings. Country GDP is simply the sum of output in each city $Y = \Sigma_d^D Y_d$. 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. As in Ahlfeldt et al. (2015), we envision that if there were to be multiple equilibria, we would select the counterfactual equilibrium closest to the observed real-world outcome.[23]

6 Estimation of Model Parameters

Section 4.1 confirmed two important model results. As (exogenous) air pollution increases, skilled labor emigrates, raising the returns to skill. Yet, our model makes clear that the observed empirical relationship between exogenous air quality $\bar{Z}_d$ and the supply of workers by skill is not simply the partial migration response to pollution due to γ_s (determining the compensating differential). Instead, in GE it is the result of corresponding migration changes as wages change, given η_s (labor supply elasticity based on local ties and preferences). The wage changes in turn depend on σ_E, the elasticity of substitution across skills in production (relative labor demand elasticity). And worker relocation further changes pollution (given ψ_1 and ψ_2), and other factors (house prices, agglomeration, etc.). As such, any empirical relationship between population changes and pollution, identify a coefficient that is a joint function of these model parameters, which in turn determine the quantitative consequences of pollution on productivity.

We estimate the elasticities: $\{\sigma_E, \eta_s, \gamma_s, \psi_1, \psi_2, \psi_3, \psi_4, \phi_1, \phi_2\}$ and city-level parameters: $\{\theta_{sd}, \alpha_{sd}, M_{sod}\}$ based on city-level relationships for a set of large and medium-sized cities where we have consistent data on all the variables across years. We control for city characteristics as before, and show robustness to alternate sources of variation. Our estimates are similar to the literature. While we could calibrate our model from the literature, we prefer using parameters causally estimated within our own model and data.

6.1 Labor Demand Curve: Estimating σ_E

Since σ_E determines the change in relative skill-unskill wages in response to changes in relative skill-unskill workers, we derive a relative demand curve from equation 4, where within city d, output (and other city characteristics) are differenced out:

$$log\frac{w_{hd}}{w_{ud}} = log\frac{\theta_{hd}}{\theta_{ud}} - \frac{1}{\sigma_E} log\frac{L_{hd}}{L_{ud}} \tag{12}$$

[23] We describe conditions that determine the existence and uniqueness of the equilibrium in Appendix E.3.

The parameter σ_E, can be estimated from this relative labor demand curve, as exogenous shifts in relative labor supply $log\frac{L_{hd}}{L_{ud}}$ 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 leverage exogenous variation in pollution to identify this relationship. For example, excess pollution from thermal inversions shifts labor supply, and traces out the labor demand curve. To estimate equation 12, we derive variation in $log\frac{L_{hd}}{L_{ud}}$ from equation 13, again instrumenting for pollution levels:

$$log\frac{L_{hd}}{L_{ud}} = \alpha_0 + \alpha_1\ log\ PM2.5_d + \varepsilon_{1d} \tag{13}$$

In Table 3, we use each of our instruments for pollution and estimate equation 13 in the columns where our outcome is the relative stock of workers $\frac{L_{hd}}{L_{ud}}$, capturing the net migration for all types of workers (whether or not they changed *hukou* location).

Table 3: Estimating Labor Demand Elasticities

IV:	Coal+Wind IV (<500km)		Coal+Wind IV(100-500km)		No. of inversions		Inversion strength	
	Ln $\frac{w_h}{w_u}$	Ln $\frac{L_h}{L_u}$	Ln $\frac{w_h}{w_u}$	Ln $\frac{L_h}{L_u}$	Ln $\frac{w_h}{w_u}$	Ln $\frac{L_h}{L_u}$	Ln $\frac{w_h}{w_u}$	Ln $\frac{L_h}{L_u}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(PM2.5)	1.000**	-1.238**	1.496***	-1.917***	0.491*	-0.945**	0.611**	-1.174**
	(0.508)	(0.629)	(0.518)	(0.741)	(0.266)	(0.421)	(0.295)	(0.485)
Observations	130	130	130	130	130	130	130	130
City Controls	Y	Y	Y	Y	Y	Y	Y	Y
Weather	N	N	Y	Y	Y	Y	Y	Y
F-stat	20.54	20.54	18.74	18.74	68.84	68.84	45.41	45.41
σ_E	1.24		1.28		1.92		1.92	

Note: We combine population census data and CLDS data. City level regressions in 2015 using 130 cities that have non-missing skill-based wage information from CLDS. 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 distance to Shanghai, Tianjin and Shenzhen seaports. In columns 1-2, we use the wind direction IV constructed using power plants located outside a given city and located within 500 km of the city. In columns 3-4, we use the wind direction IV constructed using plower plants located 100-500 km away of a given city. Robust standard errors reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. All regressions weighted by the population in 2000. $*p < 0.10, **p < 0.05, ***p < 0.01$.

In Table 1, we had used individual-level data to show how the emigration response to pollution was larger for the high-skilled, while Table A4 showed that in-migration was similarly skill-biased. These differential migration flows affect the relative stocks of skilled and unskilled workers at the city level, as we see in Table 3. In equilibrium, cities that have (exogenously) higher level of pollution, have a lower skill ratio $\frac{L_{hd}}{L_{ud}}$.

Along with the columns of Table 3 where the outcome is $log\frac{w_h}{w_u}$, we estimate equation 12. We take the ratio of the IV relationship for quantities and wages of workers. For instance, in the first two columns, we find that the elasticity of substitution across skill levels is $\sigma_E = 1.238/1.0 = 1.238$, an estimate close to the estimates in the US (Card and Lemieux, 2001). If we were to calibrate elasticities-of-substitution from the literature it would produce similar model counterfactuals below.

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

The labor supply curve in equation 10 captures bilateral migration flows between city pairs as a function of real wages and pollution at possible destinations, and migration costs between origin-destination pairs. We first parameterize migration costs M_{sod}:

$$M_{sod} = \lambda_{1s} log\, Dist_{od} + \lambda_{2s}\left(\mathbb{1}_{Migrant_{od}} \times hukou_{sd}\right)\ , \tag{14}$$

where $log\, Dist_{od}$ is the log of the distance between cities o and d, $\mathbb{1}_{Migrant_{od}}$ is an indicator for whether $o \neq d$, which is when *hukou* restrictions can bind. The skill-specific *hukou* index $hukou_{sd}$, derived from Zhang et al. (2018), measures the ease with which either skilled or unskilled workers can move into city d.[24] The distance term captures physical and psychic costs associated with moving far away from one's origin city. The interaction of *hukou* index and migration status captures institutional migration costs. Substituting equation 14 in 10 generates an estimable equation for labor supply:

$$\begin{aligned} log\, \pi_{sod} = & -\eta_s log\overline{V_{so}} + \eta_s\left(log\, w_{sd} - \nu_s log\, hp_d\right) - \eta_s\gamma_s logZ_d + \mathbf{X}\beta_x \\ & -\eta_s\lambda_{1s} log\, Dist_{od} - \eta_s\lambda_{2s}\left(\mathbb{1}_{Migrant_{od}} \times hukou_{sd}\right) + \varepsilon_{2sod}\ , \end{aligned} \tag{15}$$

where the residual $\varepsilon_{2sod} = \left(\eta_s log\, a_{sd} + \widetilde{\xi_{sod}}\right)$ includes differences in destination city amenities a_{sd} and other idiosyncratic bilateral features $\widetilde{\xi_{sod}}$. Below, we describe how we derive amenities a_{sd} from residuals 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

[24]The index measures the difficulty of obtaining a local *hukou*. A high index (most major cities) indicates a more restrictive policy. Given the strong ties to one's *hukou* city, we define origins o to be the *hukou* city.

any differences in city level characteristics (correlated with a possibly endogenous *hukou* index), that affect migrants and non-migrants similarly. Like before, we instrument for pollution using thermal inversions.

Key parameters of interest are labor supply elasticities η_s for skilled and unskilled workers. To estimate them, we need two instruments that shift labor demand curves and trace out the supply curves. We construct shift-share instruments inspired by China's export-driven economic development and entry into the world trading system. Pierce and Schott (2016) and Autor et al. (2013) study the effects of these shifts on US manufacturing employment, but here we rely on the fact that it altered skilled and unskilled wages across China (Erten and Leight, 2021; Facchini et al., 2019).

The first instrument, the NTR gap (Pierce and Schott, 2016), relies on 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 defined to be the difference between the non-NTR tariff and the NTR tariff. Unlike Pierce and Schott (2016) who focus on effects in the US, we use this instrument to study what happens to internal migration in China as real wages change across cities following trade liberalization.

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:[25]

$$NTR\,IV_{ud} = \sum_i \frac{EX_{di}^{1997}}{\sum_j EX_{dj}^{1997}} \times (\text{nonNTR tariff}_i - \text{NTR tariff}_i) \quad (16)$$

Comprehensive details about this trade shock can be found in Khanna et al. (2020), where they show that the NTR instrument better predicts higher real unskilled wages, rather than skilled wages, possibly as industries that benefited most from tariff changes were more likely to hire unskilled labor. The underlying variation is the shock to tariff uncertainty rather than city-structure (Borusyak et al., 2022). We use this as an instrument for unskilled real wage $log\,w_{ud} - \nu_u log\,hp_d$, deflating wages by local house prices.[26]

For skilled wages we derive variation from the World Import Demand (WID) for skilled

[25]While the shift-share analysis is implicitly derived from a model of multiple sectors/industries, our simplified baseline framework only has one final good for tractability.

[26]We use yearly average data on housing rents from the Xitai Real Estate Big Data depository.

industries.[27] Following Autor et al. (2013), we use WID shocks by industry, and weight them by initial export shares to derive city-level shocks. Our industry-level shifters exclude trade flows between China and the rest of the world, so as to account for unobservables related to productivity growth in China. To create an instrument for skilled-wage in equation 17 we use the share of skill-intensive industries:[28]

$$WID_{sd} = \sum_i \frac{EX_{sdi}^{1997}}{\sum_j EX_{sdj}^{1997}} \times \left(\frac{\text{World IM}_{i,2015} - \text{World IM}_{i,2004}}{\text{World IM}_{i,2004}} \right) \tag{17}$$

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 η_s, the labor supply response to changes in wages. Our estimates in columns 1 and 3 suggest that $\eta_u = 1.012$ and $\eta_h = 1.301$. These are similar to estimates of labor supply elasticities estimated by Tombe and Zhu (2019). As such, not using the instruments, and simply calibrating our model (as is done in this literature), would produce similar counterfactuals.[29]

Equation 15 shows that the coefficients on $Log(PM2.5)_d$ equal $\eta_s\gamma_s$. The γ_s parameters capture the marginal utility of clean air, and vary by skill. This determines the compensating differential for pollution. Given our estimates of η_s, we infer $\gamma_h = 0.38$, and $\gamma_u = 0.042$. $\gamma_h > \gamma_u$ implies that the skilled are more sensitive to air quality. Comparing estimated labor supply elasticities with respect to wages and pollution levels suggests that both types of workers are far more responsive to wages than they are to pollution.

Workers also respond to migration costs. Table 4 shows that migration is less likely to occur over longer distances, and if there are more *hukou* restrictions. Skilled workers are less sensitive to distance.They are more sensitive to *hukou* restrictions, even though they face fewer restrictions. The responsiveness perhaps reflects stronger preferences for

[27] We label an industry as skill intensive if the share of skilled workers in the industry is above median in the ISIC data. We construct our measure using the Indonesian manufacturing census (Amiti and Freund, 2010), to ensure no confounding effects of using the same sample to construct our measure and estimation.

[28] Khanna et al. (2020) perform many robustness checks surrounding these instruments. Recent developments in the shift-share literature discuss additional tests, such as tests for pre-trends, baseline share correlations, and standard error corrections. Khanna et al. (2020) perform these tests, noting that we rely on the assumption that in our case, the 'shifters' are exogenous (as in Borusyak et al. (2022)).

[29] In Tables D8, D9 and D10, we calibrate labor supply elasticities using parameter estimates from Tombe and Zhu (2019). The results of our counterfactual exercises remain similar.

Table 4: Estimating Labor Supply Elasticities

IV-2SLS Labor Supply	Low Skill Workers $Log\ \pi_{uod}$		High Skill Workers $Log\ \pi_{hod}$	
	(1)	(2)	(3)	(4)
$Log(PM2.5)_d$	-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

Note: Origin-destination pair level regressions across 118 origin cities and 115 destination cities for which we have data from all sources including population Census, CLDS, *hukou* index data and trade data. The measure of $Hukou\ Index_{sd}$ varies across cities and skill level. We model distance as inverse hyperbolic sine. All regressions also control for temperature, humidity, sunshine duration, and wind speed, as before when using thermal inversions as an IV. Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. The first stage relationships are described in Table C3. $*p < 0.10, **p < 0.05, ***p < 0.01$.

access to amenities (like housing purchases, medical services and children's schooling) only obtainable via accessing local *hukou*.[30]

6.3 Measuring Amenities and Productivities $\{\theta_{sd}, \alpha_{sd}, \bar{A_d}\}$

θ_{sd} varies at the city level by the amount of skill-biased capital in each city, and is measured from data on labor shares in the wage bill and the properties of a CES function:

$$\frac{w_{hd}L_{hd}}{w_{hd}L_{hd} + w_{ud}L_{ud}} = \frac{\theta_{hd}L_{hd}^{\frac{\sigma_E - 1}{\sigma_E}}}{\theta_{hd}L_{hd}^{\frac{\sigma_E - 1}{\sigma_E}} + (1 - \theta_{hd})L_{ud}^{\frac{\sigma_E - 1}{\sigma_E}}} \tag{18}$$

[30] 2017 Migrant Population Data show 54% of high-skill workers, but only 35% of low-skill workers want local *hukou*.

We plot the city-level distribution of θ_{hd} in Figure C3. Beijing, Shanghai and large urban centers have high amounts of high-skill capital than less urbanized, less developed areas.

We derive non-pollution amenities as a residual from the labor supply curve, equation 15 (Ahlfeldt et al., 2015; Bryan and Morten, 2019). We take the residual for each skill-specific regression estimated from equation 15, and derive city-and-skill specific amenities a_{sd} from destination fixed effects.

To quantify changes to output, we create a 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 $Y_d = A_d^{\frac{1}{\rho}} \left(\frac{1-\rho}{R^*}\right)^{\frac{1-\rho}{\rho}} L_d$. Using our σ_E and θ_{sd} estimates, we can create a measure of L_d, and invert the model to derive $A_d^{\frac{1}{\rho}}$.

6.4 Estimating Agglomeration and Congestion Forces (ϕ and ψ)

We use equation 19 to study how our measure of TFP correlates with pollution as instrumented with the thermal inversions instrument for pollution.

$$\widehat{\frac{1}{\rho}\log A_d} = \log \bar{A}_d + \phi_1 \log Z_d + \phi_2 \log L_{hd} + \varepsilon_{3d} \,, \tag{19}$$

where ϕ_1 is the elasticity of pollution with aggregate TFP. Notice, A_d may capture other drivers of city-level TFP, like land, housing supply, innovation. Equation 19 also allows the number of skilled workers to directly affect the amount of TFP in a city. If there are innovation spillovers, it would be captured by the agglomeration elasticity ϕ_2.[31]

In Table 5 we leverage our PM2.5 instruments and estimate $\phi_1 = -0.082$, which lies in between the Chang et al. (2019) elasticity of -0.023 for call-center workers in China, Adhvaryu et al. (2022) elasticity of -0.03 for assembly line workers in India, Chang et al. (2016) elasticity of -0.059 for fruit packers in CA, and Kahn and Li (2019) elasticity of -0.18 for public sector workers in China. Zivin and Neidell (2012) show productivity falls for low skill agricultural workers in response to high ozone levels

We use a sudden university expansion to estimate the effects of skilled workers on TFP, to capture the agglomeration or innovation spillovers. The Chinese government instituted a policy to expand college enrollment in 1999, primarily by lowering the bar for admis-

[31] Agglomeration in this model is represented by the production of non-excludable ideas. Innovators are not compensated for ideas in wages. Instead overall output increases, benefiting all in the economy.

Table 5: Pollution, Population and TFP

	Agglomeration Forces		Congestion Forces	
	Log(TFP)	Log(TFP)	Log(PM2.5)	Log(House Prices)
	(1)	(2)	(3)	(4)
Log(PM2.5)$_d$	-0.0816	-0.0595		
	(0.256)	(0.285)		
Log(L$_{hd}$)	0.0970*	0.0964*		
	(0.0528)	(0.0519)		
Log(Population)$_d$			0.266*	0.259**
			(0.150)	(0.121)
Log(L$_{hd}$/L$_{ud}$)			-0.00781	0.423***
			(0.0786)	(0.0583)
Observations	121	121	121	121
Pollution IV	No. Thermal	Thermal Strength		
First stage F-stat	25.61	15.67	12.66	12.66

Notes: We combine population census, CLDS data, and college expansion data, consistently available for 121 cities. The first column estimates the relationship between PM2.5, number of skilled workers and TFP. We use thermal inversions as an instrument for PM2.5, and leverage the higher education expansion instrument to identify the effect of the number of skilled workers. For congestion forces we use instruments for population and the skill ratio that we describe in the text. We control for region fixed effects and distance to seaports. When using thermal inversions as an instrument, we control for weather amenities (temperature, humidity, sunshine duration and wind speed). We report Kleibergen-Paap rk Wald F statistic. The first-stage relationships are in Table C3. Robustness for house-price regressions are in Table A19. $*p < 0.10, **p < 0.05, ***p < 0.01$.

sion.[32] Che and Zhang (2018) use this policy as an instrument for firm productivity, and discuss identification concerns at length. We use this same instrument to predict variation in the number of skilled workers across cities. Figure C2 describes this event. Importantly, it shows no differential pre-trends in professors and college students (who subsequently make up the skilled workforce) in cities that subsequently benefited from college expansion. Furthermore, this instrument is not correlated with the pollution instruments either.

We create measures of the number of college graduates by city and year, and use the change in graduates from 2001-05 (cohorts just before and after the university expansion policy) as our instrument.[33] We find $\phi_2 = 0.097$ when estimating equation 19, which suggests meaningful agglomeration (Moretti, 2004; Peri et al., 2015).

[32] The Ministry of Education (MOE) expanded admissions by more than 40% in 1999-2000, and by about 20% over the next five years. The enrollment rate increased from 9.8% in 1998 to 24.2% in 2009. The year 2003 saw the first flow of graduates into the job market: a 46.2% increase from the previous year.

[33] The instrument is number of college graduates in city d in 2005 minus the number of college graduates in city d in 2001. Figure C2 describes the dynamics of the expansion policy, and pre-trends.

Our model allows for the spatial distribution of population to affect pollution (equation 5). We need estimates for how changes in the number of workers and skill share in each city affect pollution and house prices. Estimating ψ_1 requires variation in the skill-ratio not driven by air quality. We leverage the college expansion policy (per baseline unskilled population in 2000) to estimate the effect of changes in skill ratios.

Simultaneously, we estimate how changes in total population affect air quality, captured by ψ_2. We leverage exogenous push factors from the cities of origin that each destination is linked to through migrant-networks at baseline. We instrument for population in each destination d using the growth in emigration between 2005-15 from all provinces, weighted by the share of migrants from each province that came to d in 2000:[34]

$$Population\ IV_{d,2015} = \sum_p \left(\frac{Migrants_{dp,2000}}{\sum_d Migrants_{dp,2000}} \times \left(\sum_{d' \neq d} Migrants_{d'p,2005-2015} \right) \right) \quad (20)$$

Note that this is different from the instruments in Card (2001); Jaeger et al. (2018). As we have rich data on outflows of migrants from provinces, we can leverage 'push factors' from sending regions, and completely omit information on inflows into destinations. The advantage is that identification relies on forces driving the outflows from provinces to all other cities, and not associated with destination d.[35]

Table 5 describes the effect of population and skill ratio on pollution. The first stage of the 2SLS appears to be strong (as in Table C3). Our two-staged least squares estimates tell us that ψ_1 is indistinguishable from zero, and $\psi_2 = 0.266$. Larger population increases the amount of pollution in a city, but the skill composition of workers has no detectable effect.

Finally, we consider how house prices may affect our predictions. We estimate the elasticity of house prices with respect to population ($\psi_3 = 0.259$),[36] and with respect to the skill ratio ($\psi_4 = 0.423$), leveraging the same instruments. A larger population raises house prices, and these prices rise substantially more with an influx of skilled workers. Robustness checks are in Table A19.

[34]When calculating the amount of out-migration between 2005 and 2015, we exclude flows to d, so as to not capture labor demand changes at the destination.

[35]We combine this idea with the strength of the Card (2001) framework, in which sources of migration at baseline determine migrant networks which attract migrants into city d after 2005, whenever there are larger outflows from the origin provinces that d was connected to. Before 2000 internal migration was highly regulated (Kinnan et al., 2018). After 2000, migration costs have fallen substantially (Tombe and Zhu, 2019).

[36]This elasticity is similar to Liang et al. (2020) who find an elasticiy of 0.21 for coastal Chinese cities.

Table 6: Summary of Estimated Model Parameters

Parameter	Value	Definition	Identifying Variation
σ_E	1.24	Skill-elasticity of substitution	Pollution-driven geographic sorting
		Relative labor demand elasticity	Changes in skill ratio affect wages
η_h	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_h	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_{hd}	[0.10,0.73]	Skill-specific productivity	Skill-labor share in wage bill
ψ_1	-0.008	Pollution response to changing skill ratio	University expansion
ψ_2	0.266	Pollution response to population	Out-migration from origin provinces
ψ_3	0.259	House price response to population	Out-migration from origin provinces
ψ_4	0.423	House price response to skill ratio	University expansion
ϕ_1	-0.0816	TFP response to pollution	Pollution IV affect TFP residual
ϕ_2	0.0970	TFP response to skilled workers	University expansion
λ_h	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 C3. 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.

6.5 Model Solution and Validation Exercises

Table 6 summarizes parameter values and sources of variation. In addition, we calibrate the housing expenditure shares $v_s = 0.159$ and $v_u = 0.182$ from the China Urban Household Survey. Output in city d depends on the set of parameters: $\{\theta_{sd}, \sigma_E, \eta, \phi_1, \phi_2, \psi_1, \psi_2, \psi_3, \psi_4\}$, a set of 'endogenous' quantities: $\{Y_d, A_d, L_{hd}, L_{ud}, K_d, Z_d\}$, and 'exogenous' quantities: $\{\bar{A_d}, \bar{Z_d}, M_{sod}, \bar{L_u}, \bar{L_h}\}$. Prices, $\{w_{hd}, w_{ud}, hp_d\}$ are determined in equilibrium, with the output being the numeraire. R^* is exogenous to the framework. Changes in exogenous pollution $\bar{Z_d}$ will affect the location of workers and TFP, thereby changing output Y_d in this city, and in other cities. Given our estimated set of parameters and exogenous quantities, we create model-predicted measures of endogenous quantities, like GDP and wages.

We solve the model starting with a list of parameters and exogenous quantities, and a set of initial conditions for endogenous variables. After estimating the parameters and exogenous quantities, we no longer use data to solve the model. The market clearing condition is the labor market equilibrium. We iteratively vary the starting value by 20% for each endogenous variable. The model converges to the same unique equilibrium.[37] Our

[37]This does not necessarily imply the equilibrium is globally unique. The existence of multiple equilibria often depends on the relative strength of agglomeration and congestion forces (Allen et al., 2020). Given the

algorithm clears the labor market. If there is an excess supply of labor in a city, wages fall, and our model converges to an equilibrium.

We test for model fit in Figure C4, and show that our model's predictions match endogenous quantities in the data. We test model-fit for predicted $Log(GDP)$, skill-premia $Log(w_h/w_u)$, $Log(L_h/L_u)$ and L_h. These are not necessarily out-of-sample tests as we use the data when estimating the model parameters, but not when solving for equilibrium which includes residuals. Across all measures, our model replicates the major spatial patterns. In Figure C5 we perform out-of-sample tests using "new" data on distance of migrant flows, and in Figure C6 we validate our measures of TFP and amenities by showing strong correlations with independent measures (for TFP we document correlations with patenting, R&D expenditure and workers, and universities).

7 Counterfactuals: The Gains from Relocating Pollution

There are three ways through which pollution result in productivity losses. First, and the one this paper seeks to highlight, skilled workers leave cities where they would be more productive. Further, since the (complementary) unskilled do not leave with the skilled, this creates a mismatch, further reducing aggregate productivity. Second, pollution reduces the agglomeration of skilled workers in productive cities. Third, pollution directly affects workers' health and lowers productivity.

In this section, we use our estimated model to conduct counterfactual exercises to quantify how large an effect pollution control policies would have on productivity via each of these three mechanisms. To isolate the health channel, we prohibit workers from changing location when pollution levels change: Without worker mobility, there is no sorting/complementarity effects nor agglomeration effects. To shut down agglomeration, we set $\phi_2 = 0$ in the TFP relationship.

There are two types of pollution control policies we consider in our counterfactuals. First, we change only the steady-state level of pollution Z_d in a city. This is similar to policies where cities are assigned explicit pollution targets they must meet, regardless of how (investing in greener technology or industrial scrubbers). In the second policy, we relocate the exogenous component of pollution only, $\bar{Z}_d$. This is similar to relocating coal-fired plants from up- to downwind regions, or targeting green investments to upwind plants.

meaningful congestion forces here, we may expect a unique equilibrium.

Relocating plants would induce new residents to move in, change production, and a revised pollution level will emerge that will not necessarily equal the initial change in $\bar{Z}_d$.

Relocating plants away from high-capital cities to low-capital cities, will benefit certain cities but hurt others. Even though the relocation may improve aggregate productivity, the distributional effects are more complex.[38] These policies may not be costless: our quantified benefits inform policy given any known menu of costs.

7.1 Changing Pollution in One City: Beijing

Table 7 describes the effects of changes to pollution and migration policy in a highly productive and polluted city, Beijing. We first reduce steady state PM2.5 by 50%, a policy akin to setting a pollution cap for Beijing. This raises GDP per worker in Beijing by 14.4%. The health channel raises productivity by 5.8%. The pure relocation channel (not accounting for agglomeration) raises incomes by 4.5%. If we allow for skilled workers to drive agglomeration spillovers, GDP per worker due to relocation would rise by 8.1%. Productivity gains through the indirect spatial sorting channel are larger than the direct health benefits of air quality improvement. The overall improvements to GDP are not merely the sum of the channels, as they meaningfully interact with each other.

Second, we examine the effects of reducing the *exogenous component* of PM2.5, allowing steady state values of pollution to adjust when city population and skill composition change. This policy is similar to relocating a power plant upwind of Beijing to elsewhere. Here we obtain similar, albeit slightly smaller, effects on GDP per worker and wages. The mildly muted effects reflect the fact that when exogenous pollution is reduced, an influx of workers may increase pollution and mitigate some gains. Again, relocation effects are larger in magnitude than the direct health benefits.

In the third and fourth rows of Table 7, we examine the effects of relaxing *hukou* restrictions in Beijing by 50%, but holding pollution fixed. When we relax the skilled *hukou*, GDP per worker rises by 8.2%, more than half of which is driven by the relocation channel, and the rest by agglomeration. Lowering unskilled *hukou* restrictions lowers GDP per capita, through a compositional change in population – there are now more low-wage work-

[38] While we study outcomes in equilibrium steady states, the movement of individuals may evolve dynamically over time (Heblich et al., 2021). To be consistent with our steady-state approach, our primary source of pollution variation relies on mostly cross-sectional variation (based on the placement of power plants). Yet, our results are consistent when using more dynamic variation derived from thermal inversions over time.

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

	Change in GDP per Worker in Beijing (%)			
	Overall Change	Health Channel	Relocation	Relocation+Agglom
	(1)	(2)	(3)	(4)
Reduce steady state PM2.5	14.370	5.819	4.480	8.080
Reduce exogenous part of PM2.5	12.773	5.344	3.777	7.052
Relax skilled hukou	8.174	0.000	5.109	8.174
Relax unskilled hukou	-3.775	0.000	-4.686	-3.775
Reduce PM2.5 & relax skilled hukou	22.959	5.819	9.147	16.197
Reduce PM2.5 & relax unskilled hukou	10.533	5.819	0.036	4.454

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 lower *hukou* restrictions for each skill level (rows 3 and 4) by 50%, keeping pollution fixed. Rows 5 and 6 lower the *hukou* regulations by 50% 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 shows the pure relocation channel, and Column 4 also incorporates agglomeration as a consequence of relocation. Table D1 shows effects on nearby cities, and Table D1 shows effects of the simplest model without any externalities.

ers in Beijing. The last two rows combine the changes in steady-state pollution and *hukou* restrictions. Combining pollution abatement and relaxing the skilled *hukou*, GDP increases by 9.2% due to reallocation of workers, and 16.2% when including the agglomeration of skilled work. These are much larger than the direct health benefits of clean air, 5.8%.[39]

Relocation affects productivity and wages in the model through multiple channels. First, immigration changes the skill *composition* of the city population. Average incomes rise if skilled workers move in. Second, the effect of skilled workers on city productivity is larger when the city (like Beijing) has a lot of skill-biased capital. Third, inflow of skilled workers lowers the skilled wage due to a labor supply effect. However, that raises unskilled wages because the unskilled are estimated to be fairly complementary in production. Finally, agglomeration forces will raise average incomes for all skill groups. When *hukou* relaxation is combined with pollution reduction, the consequent skilled immigration raises average incomes substantially.

We move beyond overall GDP per capita to examine distributional consequences on the wages of each skill group in Table D2. When we reduce the steady state amount of pollution in row 1, skilled wages rise slightly. The improved productivity of skilled workers from

[39]Panel A in Appendix Table D1 examines changes to GDP per worker, focusing solely on the main relocation effects and shutting down agglomeration or congestion effects.

reduced 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 18.3%. Most of this is driven by the relocation channel: When skilled workers enter Beijing, complementary unskilled workers become more productive. As a consequence, average wages in the city rise by 14.4% (i.e., the increase we saw in GDP per worker in Table 7). Reducing the exogenous part of pollution (row 2) produces similar distributional effects.

When we relax the *hukou* restrictions in rows 3 and 4 in each skill group, we see that allowing in more workers of a particular skill group lower the wages of that group, while raising the productivity of the other (complementary) worker type. When we combine relaxing *hukou* regulations with changes to PM2.5, there are more positive effects on wages, especially for the unskilled. Panel B in Table D1 shows effects on nearby cities.

7.2 Relocating Pollution Away from Skill-biased Capital

Table 8 considers a different type of 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_{hd}) to regions with less skill-biased capital. This could, for instance, entail relocating coal-fired plants away from technology hubs, financial centers, and nodes of professional service activities. Spatial reorientation of coal processing is perhaps more feasible than reducing nationwide production and pollution in the aggregate. In the first row, we relocate steady state pollution by setting pollution caps based on the amount of skill-biased capital in the city. Overall GDP in the country increases by 6.7%, and a substantial proportion of this increase is driven by the relocation of workers. The contribution of the health channel is a 2.6% increase in GDP, while the relocation channel alone raises GDP by 2.2%. Agglomeration plays a minor role.

When we relocate the exogenous part of pollution (say, shift an upwind plant away from a productive city, to a less productive one), the GDP increase is smaller. As people relocate to the more productive cities, pollution levels again rise through the feedback loop in the model (where pollution is a function of population), and dampens the benefits. The contribution of the relocation channel is slightly larger than the health channel.[40]

As a benchmark, in row 3, we relax *hukou* restrictions in the top-tier cities by 50%. This raises GDP by less than the pollution changes, but, by construction, is solely driven

[40]The health channel playing a positive role suggests that sources of pollution are concentrated in more populous cities. Moving pollution away from large cities can improve population-weighted average health.

Table 8: The Productivity Effect of Relocating Pollution Across Cities

	Change in GDP per Worker in China (%)			
	Overall changes	Health	Relocation	Relocate+Agglom
	(1)	(2)	(3)	(4)
Relocate steady state PM2.5	6.702	2.604	2.205	2.277
Relocate exogenous part of PM2.5	3.670	1.484	1.563	1.615
Relax hukou	2.585	0.000	2.340	2.585
Relax overall mobility constraints	6.968	0.000	6.421	6.968
Relocate PM2.5 & relax hukou	8.329	2.604	3.615	3.844
Relocate PM2.5 & lower migration costs	13.832	2.604	8.156	8.814

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 row 3, we relax *hukou* restrictions in the 24 top tier cities by 50%, keeping pollution fixed. In row 4 we relax overall migration costs to the 24 top tier cities by 50%, keeping pollution fixed. Column 1 shows overall gains to GDP. Column 2 shows the GDP increase as a consequence of health effects only. Column 3 shows gains due to the re-allocation of labor only. Column 4 shows the gain to GDP accounting for changes in TFP due to changes in re-location and the agglomeration of skilled workers. Table D3 shows results for the simplest model focusing on relocation, and no externalities.

by worker relocation. When we lower overall migration costs in these top-tier cities (row 4) by 50%, the increase in GDP is similar to relocating steady state pollution (about 6.97%). Lowering overall migration costs can be thought of as a policy mix of relaxing *hukou* restrictions and building more transportation infrastructure to connect cities. Combining reductions to mobility costs to top-tier cities with relocating pollution produces much larger effects on GDP per worker.[41].

The major lessons are: (a) A spatial reallocation of pollution away from cities that have greater potential for skilled productivity (those with most skill-biased capital) can raise national income as skilled workers relocate to where they are most productive, (b) Pollution caps produce larger productivity effects than relocating sources of pollution, as worker re-sorting undermines benefits, and (c) Combining relaxations to mobility restrictions with pollution reductions produce large income gains.

Table D4 shows wage effects. Relocating pollution away from cities with more skill-biased capital raises skilled wages, but has little effect on unskilled wages. This is a consequence of the baseline distribution of skill groups across cities that see pollution changes.

[41] In Appendix Table D3, 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. The results are qualitatively similar to Table 8

As skilled workers relocate to cities with more skill-biased capital, their overall productivity increases, as there is now a better matching of workers to capital. Skilled workers who were already resident in such cities may see a dampening of their wages, but could also benefit from agglomeration economies. On net, we find that skilled wages rise by 17.7%, and the relocation channel raises skilled wages by 10.5%. The health channel also raises skilled wages, as skilled workers already tend to locate in high skill-biased capital cities which now saw a reduction in pollution.

Relaxing *hukou* restrictions and lower migration costs (rows 3 and 4) raise wages for both skilled and unskilled, as workers match better with where their marginal products are higher. The combination of lowering migration costs and relocating pollution (row 6) raises skilled wages by as much as 27.8% and unskilled wages by 6.7%, almost entirely due to changes in internal migration patterns.

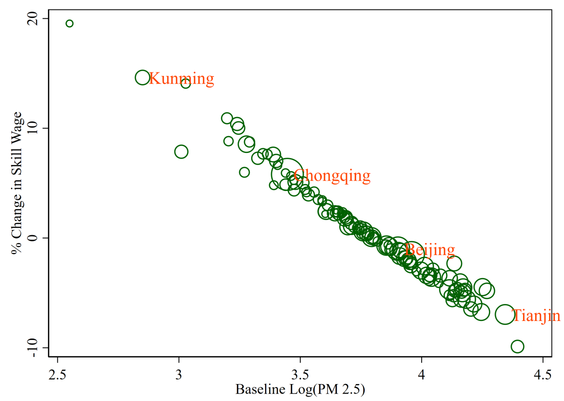
However, wage effects do not capture the entirety of the welfare consequences, as pollution and migration costs also determine welfare. This is particularly important to acknowledge, as relocating pollution to less productive areas may be undesirable from an environmental justice point of view, if it makes unskilled workers in poor cities worse off. Table D5 examines changes to welfare by skill (as defined in equation E.6). Relocating pollution away from cities with skill-biased capital raises the welfare of skilled workers by 29.7%, as it raises their wages and lowers their pollution experience as most skilled workers are already located in such cities. Unskilled workers, however, see modest improvements in overall welfare. National welfare improves by 4.6%.

Lowering migration costs (row 4) raises the welfare of both the skilled and unskilled by 17.5% and 10.1% respectively. The combination of relocating pollution and lowering migration costs improves country welfare by 16.9%. The overall changes in welfare are not simply the sum of the two counterfactuals, highlighting the interplay between migration restrictions and pollution exposure. Relaxing migration restrictions improves welfare by lowering mobility costs and allowing access to high wages; but when combined with lower pollution, the in-migrants also benefit from air quality.

The numbers in Table D5 make clear that while relocating pollution is sensible on efficiency grounds, they raise welfare by benefiting skilled workers. Unskilled workers are also better off, but their gains are only substantial when reducing overall mobility costs.

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

Figure 8: Explaining the Wage Gap with Worker Relocation



Notes: We plot the change in the skilled wage, solely due to changes in worker location (the relocation channel only), when the amount of pollution in the city is changed to be that 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 pollution is equalized across cities. The size of the bubbles represent the baseline population in 2000.

A persistent puzzle animating a large literature in development and macroeconomics is that despite existent productivity gaps across regions, worker mobility does not equalize wages. We use our estimated model to explore: how much of the cross-city wage gaps is 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 total country's pollution the same as before. This means raising pollution levels in low pollution and low skill-biased-capital cities, and lowering them in polluted, productive cities. In Figure 8, we show how wages change, to quantify how much of the wage gap across cities is due to existing patterns of pollution. For instance, Tianjin and Chongqing are two comparable representative provincial-level cities: As pollution is lowered in Tianjin, there is an inflow of workers that lowers wages. The change in wage is the same as the (endogenously determined) compensating differential. Conversely, wages rise in Chongqing as workers emigrate out. On net, the skilled wage gap between Tianjin and Chongqing is bridged by 18%, due only to the pollution-induced reallocation channel. Once we incorporate the health and agglomeration benefits of reducing pollution, a smaller, 14.4% of the gap gets bridged.

7.4 Consequences of the 2013 City-level Pollution Caps

Finally, we quantify the productivity implications of an actual pollution control policy recently implemented by the Chinese government. On Sep 10, 2013, the State Council of China issued an *Air Pollution Prevention and Control Plan*, which states, "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 fall by around 25%, 20% and 15% respectively. PM2.5 concentration in Beijing shall be controlled below 60 $\mu g/m^3$." This plan sensibly targeted China's three most productive areas.

Table 9: The Productivity Effect of Pollution Regulation

	Change in GDP per Worker in China (%)			
	Overall changes	Health	Relocation	Relocate+Agglom
	(1)	(2)	(3)	(4)
Control PM2.5	3.570	1.519	1.702	1.937
Control PM2.5 & relax hukou	6.300	1.519	4.064	4.570
Control PM2.5 & lower migration costs	10.773	1.519	8.157	8.974

Notes: In this counterfactual 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 high-tier cities (row 2), and overall migration costs (row 3) by 50%. Column 1 shows the gain to country GDP. Column 2 shows the GDP gain from the health-productivity channel. Column 3 shows the GDP change from relocation, and Column 4 also accounts for agglomeration. We examine how this policy affects wages of workers in Table D6 and their welfare in Table D7.

As shown in Table 9, we predict this policy – targeted at only a subset of cities – to increase country-level GDP by 3.6%, mostly driven by workers relocating to more productive cities. While this is already impressive, our model further suggests that if China were to lower migration costs to allow its citizens to take full advantage of the new pollution controls in productive cities, GDP per worker would rise even more, by as much as 10.8%.

Our analysis suggests that while pollution control is important, ignoring the spatial resorting effects of pollution control leaves large bills on the sidewalk. Given how sensitive (skilled) Chinese citizens are to pollution and migration costs, our model clarifies that pairing pollution control with easing *hukou* restrictions could produce large benefits to society. Table D7 shows that a combined policy of both pollution control and easing mobility would also make the welfare gains more equitable (as in equation E.6). The 2013 pollution-caps policy raised skilled welfare by 7.2%, but have little effect on the unskilled, who are less pollution-sensitive. However, lowering migration costs as a complementary policy would improve welfare for both the skilled and unskilled.

8 Conclusion

Our analysis highlights the macroeconomic consequences of an important new pattern of mobility, and proposes a new channel through which pollution adversely affects aggregate productivity. As economies grow and industrial activity pollutes the environment, workers – especially those who are more educated and skilled – emigrate in search of better air quality. Not only is this costly for the polluted cities that skilled workers leave, this process lowers productivity and aggregate economic growth by creating a spatial mismatch between skilled and unskilled workers, and by inducing skilled workers to move out of areas where they would contribute more to the economy. Other work documents that pollution lowers productivity by making workers unhealthy (Adhvaryu et al., 2022; Kahn and Li, 2019; Zivin and Neidell, 2012), and our contribution is to quantify the productivity losses stemming from differential mobility of skilled workers in response to pollution, which we find to be just as important as the pollution-health effects. We further document that mobility costs (both physical, and via *hukou* policy) exacerbate these economic losses, and that migration and pollution-control policies are interlinked. This evidence directly speaks to tensions between environmental regulation and urbanization in the developing world (Balboni, 2021; Glaeser, 2014b).

Finally, our analysis sheds light on an important puzzle in the development and macroeconomics literature: Why are there large productivity gaps across regions within countries (Bryan and Morten, 2019; Gollin et al., 2014), and why don't workers move to arbitrage the gaps (Bryan et al., 2014; Heise and Porzio, 2021)? Understanding factors that prevent efficient allocations of inputs is consequential for our understanding of aggregate productivity and growth (Hsieh and Klenow, 2009). We find that distaste for pollution accounts for 14% of the wage difference across representative pairs of cities. While most commentators fear that pollution control policies hurt firm performance, we show that such policies has the potential to bring about large productivity gains in China.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2022). Management and Shocks to Worker Productivity. *Journal of Political Economy 130*(1), 1–47.

AFP News (2019, March). Asia's Pollution Exodus: Firms Struggle to Woo Top Talent. *Agence-France Press News Agency*.

Ahlfeldt, G., S. Redding, D. Sturm, and N. Wolf (2015). The Economics of Density: Evidence from the Berlin Wall. *Econometrica 83*(6), 2127–2189.

Allen, T., C. Arkolakis, and Y. Takahashi (2020). Universal Gravity. *Journal 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.

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.

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.

Bai, X., K. Krishna, and H. Ma (2017). How you export matters: Export mode, learning and productivity in china. *Journal of International Economics 104*, 122–137.

Balboni, C. (2021). In Harm's Way: Infrastructure Investments and the Persistence of Coastal Cities. *Working Paper*.

Barwick, P. J., S. Li, L. Lin, and E. Zou (2019). From Fog to Smog: The Value of Pollution Information. *National Bureau of Economic Research, WP 26541*.

Bassi, V., M. Kahn, N. Lozana Gracia, T. Porzio, and J. Sorin (2021). Pollution in Ugandan Cities: Do Managers Avoid it or Adapt in Place? *Working Paper*.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

Brandt, L., C.-T. Hsieh, and X. Zhu (2008). Growth and Structural Transformation in China. *China's Great Economic Transformation*, 683–728.

Brandt, L., T. Tombe, and X. Zhu (2013). Factor Market Distortions Across Time, Space and Sectors in China. *Review of Economic Dynamics 16*(1), 39–58.

Brown, J., T. Hossain, and J. Morgan (2010). Shrouded Attributes and Information Suppression: Evidence from the Field. *The Quarterly Journal of Economics 125*(2).

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica 82*(5).

Bryan, G. and M. Morten (2019). The Aggregate Productivity Eeffects of Internal Migration: Evidence from Indonesia. *Journal of Political Economy 127*(5), 2229–2268.

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. Graff Zivin, T. Gross, and M. Neidell (2016). Particulate Pollution and the Productivity of Pear Packers. *American Economic Journal: Economic Policy 8*(3), 141–169.

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).

Chen, S., P. Oliva, and P. Zhang (2022). The Effect of Air Pollution on Migration: Evidence from China. *Journal of Development Economics 156*, 102833.

Chen, Y. (2023). Environmental Regulation, Local Labor Market, and Skill Heterogeneity. *Regional Science and Urban Economics 101*, 103898.

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.

Cleary, S. (1999). The Relationship between Firm Investment and Financial Status. *The Journal of Finance 54*(2), 673–692.

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).

Combes, P.-P., S. Demurger, S. Li, and J. Wang (2019). Unequal Migration and Urbanisation Gains in China. *Journal of Development Economics 142*(1), 102328.

Dahl, G. (2002). Mobility and the Retun to Education: Testing a Roy Model with Multiple Markets. *Econometrica 70*(6), 2367–2420.

Dell, M., B. F. Jones, and B. A. Olken (2012). Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics 4*(3), 66–95.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review 106*(3), 479–524.

Dingel, J. I., A. Miscio, and D. R. Davis (2019). Cities, Lights, and Skills in Developing Economies. *Journal of Urban Economics*, 103174.

Duranton, G. and D. Puga (2005). From Sectoral to Functional Urban Specialisation. *Journal of Urban Economics 57*(2), 343–406.

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).

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), 10384–10389.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial Sorting. *Journal of Political Economy 122*(3), 554–620.

Erten, B. and J. Leight (2021). Exporting Out of Agriculture: The Impact of Wto Accession on Structural Transformation in China. *Review of Economics and Statistics 103*(2).

Facchini, G., M. Y. Liu, A. M. Mayda, and M. Zhou (2019). China's Great Migration: The Impact of the Reduction in Trade Policy Uncertainty. *Journal of International Economics 120*, 126–144.

Feenstra, R., H. Deng, C. Hong, P. Luck, A. Ma, H. Ma, S. Yao, G. Wright, and M. Xu (2018). Chinese and hong kong international trade data. *Center for International Data, University of California, Davis*.

Ferrini, F. (1979). Thermodynamic Theory of Temperature Inversion in the Earth Atmosphere. *Astrophysics and Space Science 61*(2), 267–275.

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

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), 131–164.

Glaeser, E. (2014a). A World of Cities: The Causes and Consequences of Urbanization in Poorer Countries. *Journal of the European Economic Association 12*(5), 1154–1199.

Glaeser, E. (2014b). The Supply of Environmentalism: Psychological Interventions and Economics. *Review of Environmental Economics and Policy 8*(2), 208–229.

Glaeser, E. and M. E. Kahn (2010). The Greenness of Cities: Carbon Dioxide Emissions and Urban Development. *Journal of Urban Economics 67*(3), 404–418.

Gollin, D., D. Lagakos, and M. Waugh (2014). The Agricultural Productivity Gap. *Quarterly Journal of Economics 129*(2), 939–993.

Greenstone, M., G. He, R. Jia, and T. Liu (2020). Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution. *American Economic Review: Insights*.

Hanlon, W. (2020). Coal Smoke, City Growth, and the Costs of the Industrial Revolution. *The Economic Journal 130*(626), 462–488.

He, G., M. Fan, and M. Zhou (2016). The Effect of Air Pollution on Mortality in China: Evidence from the 2008 Beijing Olympic Games. *Journal of Environmental Economics and Management 79*.

He, G., T. Lio, and M. Zhou (2020, June). Straw Burning, PM2.5, and Death: Evidence from China. *Journal of Development Economics 145*(102468).

He, G., Y. Xie, and B. Zhang (2020, June). Expressways, GDP, and the Environment: The Case of China. *Journal of Development Economics 145*(102485).

Heblich, S., A. Trew, and Y. Zylberberg (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy 129*(5), 1508–1552.

Heise, S. and T. Porzio (2021). The Aggregate and Distributional Effects of Spatial Frictions. *National Bureau of Economic Research, WP 28792*.

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 124*(4), 1403–1448. November.

Hsieh, C. T. and E. Moretti (2018). Housing Constraints and Spatial Misallocation. *American Economic Journal: Macroeconomics 11*(2), 1–39.

Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg (2022). Migrants and firms: Evidence from china. *American Economic Review 112*(6), 1885–1914.

Ito, K. and S. Zhang (2019). Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Journal of Political Economy 128*(5), 1627–1672.

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 J. P. Nilsson (2018). Economic Status, Air Quality, and Child Health: Evidence from Inversion Episodes. *Journal of Health Economics 61*, 220–232.

Kahn, M. E. and P. Li (2019). The Effect of Pollution and Heat on High Skill Public Sector Worker Productivity in China. *National Burea of Economic Research, WP 25594*.

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

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig (2010). Did Securitization Lead to Lax Screening? Evidence from Subprime Loans. *The Quarterly journal of economics 125*(1).

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.

Liang, W., R. Song, and C. Timmins (2020). Frictional Sorting. *National Bureau of Economic Research, WP27643*.

Molina, T. (2021). Pollution,Ability, and Gender-Specific Investment Responses to Shocks. *Journal of the European Economic Association 19*(1), 580–619.

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-level Production Functions. *American Economic Review 94*(3), 656–690.

Moretti, E. (2011). Local Labor Markets. *Handbook of Labor Economics 4b*. Elsevier North Holland, Amsterdam.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics 122*(2), 569–600.

Peri, G., K. Shih, and C. Sparber (2015). STEM Workers, H-1B Visas, and Productivity in US Cities. *Journal of Labor Economics 33*(S1), S225 – S255.

Pierce, J. R. and P. K. Schott (2016). The Surprisingly Swift Decline of U.S. Manufacturing Employment. *American Economic Review 106*(7), 1632–62.

Piyapromdee, S. (2021). The Impact of Immigration on Wages, Internal Migration, and Welfare. *The Review of Economic Studies 88*(1), 406–453.

Rendon, A., J. Salazar, C. Palacio, V. Wirth, and B. Brotz (2014). Effects of Urbanization on the Temperature Inversion Breakup in a Mountain Valley with Implications for Air Quality. *Journal of Applied Metereology and Climatology 53*(4), 840–858.

Sharan, M. and S. Gopalakrishnan (2003). Mathematical Modeling of Diffusion and Transport of Pollutants in the Atmospheric Boundary Layer. *Pure and Applied Geophysics 160*(1), 357–394.

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.

Xue, S., B. Zhang, and X. Zhao (2021). Brain drain: The Impact of Air Pollution on Firm Performance. *Journal of Environmental Economics and Management 110*, 102546.

Yu, B., W.-S. Lee, and S. Rafiq (2020). Air Pollution Quotas and the Dynamics of Internal Skilled Migration in Chinese Cities. *Working Paper*.

Yu, M. (2015). Processing Trade, Tariff Reductions and Firm Productivity: Evidence from Chinese Firms. *The Economic Journal 125*(585), 943–988.

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), 3652–73.

Online Appendix

Table of Contents

A Robustness of Empirical Specifications **A1**

A.1 The Different Instruments and their Sources of Variation A1

A.2 Alternative Model Specifications, Controls and Samples A12

B Air Pollution Data Disclosure in China **A18**

C Additional Tables and Figures **A21**

D Additional Counterfactual Results **A27**

E Additional Model Derivations **A34**

E.1 Deriving Labor Supply and Welfare A34

E.2 Elasticity of Capital, and Skill-biased Capital A36

E.3 On Existence and Uniqueness of the Equilibrium A37

F Data Appendix **A39**

A 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. We explore threats to identification, different instruments, alternate model specifications and data.

A.1 The Different Instruments and their Sources of Variation

Here we explore threats to identification for our 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 a given city and coal consumption – to determine where to place new plants. We thus exclude any plants built within different distance radii around the city. We may still think that *newly* built plants are endogenously placed. Yet, our results are robust to relying on old power plants, and to make it even more conservative the cities near 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). Even though we fail to find substantial effects on out-migration rates, they do help predict differential in-migration by skill.

A.1.1 Instrumental Variable Estimates

In Table A1 we show the strength of the first stage relationships between our different instruments and our independent variable of interest.

In the first three columns of Table A2, we estimate a stacked regression analogue of Table 1, where we include an interaction term between pollution and high-skill indicators (and a control for the high-skilled indicator). Here, we instrument for pollution, and also for the interaction between pollution and high-skilled (using the interaction between the instrument and the high-skilled indicator). The results indicate that pollution increases emigration, and this effect is larger for the high-skilled.

In the next three columns of of Table A2, we look at different ways to reformulate the estimation sample. Instead of splitting up the sample into low and high skilled, we split it up into three categories: high school or below, those with some college education, and

Table A1: The First Stage Across Different Instruments

Panel A: City-level	Dependent variable: Log (PM2.5)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Wind+Coal IV	0.0240***	0.0113***	0.00956***						
	(0.00211)	(0.00235)	(0.00239)						
Number of inversions				0.00222***	0.00159***	0.00201***			
				(0.000350)	(0.000329)	(0.000293)			
Strength of inversions							0.000730***	0.000448**	0.000725***
							(0.000214)	(0.000185)	(0.000168)
Observations	332	332	332	332	332	332	332	332	332
R-squared	0.297	0.394	0.444	0.212	0.441	0.538	0.097	0.389	0.483
City Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Weather Controls	No	No	Yes	No	No	Yes	No	No	Yes
Panel B: Individual-level	Dependent variable: Log (PM2.5)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Wind+Coal IV	0.0200***	0.0136***	0.0130***						
	(0.00191)	(0.00188)	(0.00189)						
Number of inversions				0.00212***	0.00184***	0.00227***			
				(0.000285)	(0.000254)	(0.000224)			
Strength of inversions							0.000822***	0.000670***	0.000944***
							(0.000165)	(0.000151)	(0.000132)
Observations	761,548	761,548	761,548	761,548	761,548	761,548	761,548	761,548	761,548
R-squared	0.284	0.403	0.447	0.248	0.441	0.551	0.160	0.379	0.491
City Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Demographics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Weather Controls	No	No	Yes	No	No	Yes	No	No	Yes

Notes: City-level regressions of 332 cities in Panel A and individual-level regressions across 332 cities in Panel B. Panel A shows the first-stage results for city-level IV regressions and Panel B shows the first-stage results for individual-level IV regressions. Standard errors clustered at the hukou city level are reported in parentheses. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Weather controls include temperature, wind speed, sunshine duration and humidity.

those with college or above education. A steep education gradient is apparent, where the elasticity of migration with respect to PM2.5 is higher for higher levels of education.

Finally, Table A3 shows similar results using the combinations of the Wind IV and thermal inversions instruments, and the corresponding over-identification tests.

We further examine the association between pollution and *destination* choices. We regress city-level in-migration on local pollution concentration, and use our instruments to deal with the endogeneity of air quality in destination cities. In Table A4 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 not only results in the outflow of high-skilled workers but also reduce their inflow.

Table A2: Stacked Regressions and Disaggregated Skill Levels

	Leave hukou city indicator					
	(1)	(2)	(3)	(4)	(5)	(6)
	Stacked Regressions (with High-skill interaction)			Disaggregated Education Levels		
	Full sample Wind IV	Full sample Thermal Num	Full sample Thermal Strength	High school or below	Some college Wind IV	College or above
Log(PM2.5)	0.0751*	0.0845***	0.0704***	0.0609	0.116***	0.176***
	(0.0388)	(0.0203)	(0.0230)	(0.0423)	(0.0401)	(0.0409)
Log(PM2.5) × High-skill	0.0622**	0.0810**	0.110**			
	(0.0251)	(0.0376)	(0.0528)			
Observations	761,548	761,548	761,548	643,124	64,598	53,826
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	No	Yes	Yes	No	No	No

Notes: Individual level regressions across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant (in column 1 and columns 4-6), and using the number of thermal inversions (in column 2) and the strength of thermal inversions (in column 3). We also control for a high-skilled indicator in column 1-3. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. Weather controls include temperature, wind speed, sunshine duration and humidity.

Table A3: Combined Instruments

	Dependent variable: Leave hukou city indicator					
	Wind and Coal +Number of inversions			Wind and Coal +Strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)
Log (PM2.5)	0.0892***	0.0836***	0.124***	0.0795***	0.0722***	0.129***
	(0.0189)	(0.0197)	(0.0228)	(0.0206)	(0.0213)	(0.0263)
Observations	761,548	643,124	118,424	761,548	643,124	118,424
Hansen J statistic	0.063	0.407	2.162	0.03	0.041	2.194
Hansen P value	0.802	0.523	0.142	0.862	0.839	0.139
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	68.36	65.18	54.29	45.91	43.23	36.66

Notes: Individual level regressions across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. 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 temperature, wind speed, sunshine duration and humidity.

Table A4: Pollution and In-migration

Panel A	Dependent variable: Share of In-migrants								
No weather	Wind+Coal IV			Number of inversions			Strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	-1.856***	-1.853***	-2.259***	-1.204**	-1.137**	-1.645**	-1.404**	-1.275**	-2.152**
	(0.529)	(0.520)	(0.716)	(0.500)	(0.475)	(0.677)	(0.643)	(0.606)	(0.903)
Observations	329	329	329	329	329	329	329	329	329
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	No	No	No	No	No	No	No	No	No
F-test of IVs	41.91	41.91	41.91	39.56	39.56	39.56	16.36	16.36	16.36

Panel B	Dependent variable: Share of In-migrants								
Weather controls	Wind+Coal IV			Number of inversions			Strength of inversions		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	-1.833***	-1.825***	-2.266***	-1.204***	-1.181***	-1.451***	-1.359***	-1.301***	-1.794***
	(0.529)	(0.513)	(0.728)	(0.373)	(0.357)	(0.503)	(0.421)	(0.401)	(0.583)
Observations	329	329	329	329	329	329	329	329	329
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	33.81	33.81	33.81	91.64	91.64	91.64	47.63	47.63	47.63

Notes: City level regressions of 329 cities. We drop cities with missing in-migration rates. Dependent variable is the log share of in-migrants in city population. Independent variable is destination city PM2.5. Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant, and using the number of thermal inversions and the strength of thermal inversions. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. In Panel B, weather controls include temperature, wind speed, sunshine duration and humidity.

A.1.2 Endogeneity Concerns over Instrumental Strategies

In this section we test certain identification assumptions. Our first instrument is the interaction between the three components: 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, and so the instrument may be correlated with unobservable characteristics of nearby cities. In Table A5, we exclude any plants built within 200km of a given city (first two columns), and then within 400km of the city (last two columns). Our results are similar to before, with an increase in precision.

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 A6 we exclude newly built

Table A5: Different Distance Bins for Selection of Plants

	Dependent variable: Leave hukou city indicator					
	Distance 200-600km			Distance 400-800km		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)
Log (PM2.5)	0.0947***	0.0798**	0.163***	0.0973***	0.0834***	0.168***
	(0.0365)	(0.0396)	(0.0383)	(0.0304)	(0.0319)	(0.0385)
Observations	761,548	643,124	118,424	761,548	643,124	118,424
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	56.73	50.95	43.11	59.73	53.08	58.44

Notes: Individual level regressions across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification based on the interaction between wind direction, distance to coal plant, and coal consumption of power plant. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

plants from the 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.

As the IV is constructed using the interaction between wind direction, distance to coal plants and the capacity of plants, it is natural to ask which of the three components drives our results. In Table A7, we control for the distance between each city and power plants with little affect on our empirical pattern. In Table A8 we try different versions of the instrument in which we exclude each of the three components, respectively. Our results hardly change when we exclude distance to coal plant and coal consumption. In contrast, the coefficient estimates of air pollution become meaningfully smaller when we exclude the component of wind direction, indicating that our main IV results are primarily driven by the variation in wind direction across locations. Wind direction is determined by nature and is stable over long periods of time, thus it can be considered as exogenous to local economies.

One concern with the exogeneity of wind direction is that the Chinese government might select thermal plant locations in a way that pollution did not travel to populated or politically important cities. If that were the case, coal-fired plants are less likely to be located upwind of such cities. In Table A9, we present the number of large-scale thermal

Table A6: 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
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (PM2.5)	0.0611	0.138***	0.0761*	0.146***	0.0762*	0.150***	0.0992**	0.156***
	(0.0421)	(0.0379)	(0.0451)	(0.0412)	(0.0432)	(0.0401)	(0.0469)	(0.0398)
Observations	643,124	118,424	643,124	118,424	643,124	118,424	643,124	118,424
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	44.39	40.57	33.08	34.42	35.28	42.99	30.87	38.36

Notes: Individual level regressions across 332 cities. 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. We report Kleibergen-Paap rk Wald F statistic. 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 A7: Controlling for Distance to Cities

	Full sample	Low edu	High edu
	(1)	(2)	(3)
Log(PM 2.5)	0.0700*	0.0541	0.133***
	(0.0394)	(0.0429)	(0.0373)
Observations	750,822	633,288	117,534
R-squared	0.029	0.03	0.04
City Controls	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
F-Test of IVs	57.24	51	42.83

Notes: Individual level regressions across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification based on the interaction between wind direction, distance to coal plant, and coal consumption of power plant. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. We control for the average distance between each city and power plants located within 500 km of the city.

plants located upwind of five largest metropolitan areas in 2014 along with their total coal consumption. Beijing and Tianjin are among the most populated and politically important cities in Northern China. The ratio of the upwind large thermal plants to the total number of

Table A8: Decomposing the Wind Direction IV

	Dependent variable: Leave hukou city indicator								
	IV:Excluding distance			IV:Excluding coal consumption			IV:Excluding wind direction		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	0.0943**	0.0785*	0.163***	0.0905**	0.0765**	0.144***	0.0360	0.0199	0.0911***
	(0.0392)	(0.0424)	(0.0392)	(0.0358)	(0.0385)	(0.0350)	(0.0343)	(0.0378)	(0.0330)
Observations	761,548	643,124	118,424	761,548	643,124	118,424	761,548	643,124	118,424
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	46.07	40.52	39.33	79.24	73.06	63.70	55.14	50.21	38.13

Notes: Individual level regressions across 332 cities.Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. 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 distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

large thermal plants is 22.4% and 50.9% for Beijing and Tianjin, respectively. The mean of the ratio for all Chinese cities is 37.8%, which is between the values in Beijing and Tianjin. This may suggest that the Chinese government does not necessarily locate the coal-fired plants away from populated or politically important cities.

To further test whether politicians avoid populated, politically important and rich cities when building new plants, we explore whether baseline city features predict newly built plants. In Table A10, we explore whether city-level characteristics in 2004 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, major provincial capitals and coal producing regions (Table A15).

In Table A11, we construct placebo instruments, artificially changing the wind angle by 90, 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 PM2.5 or emigration.

Finally, we turn our attention to the thermal inversions IV, used extensively by researchers in many different contexts (Arceo et al., 2016; Chen et al., 2022; Hicks et al.,

Table A9: 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	22.41%	2372	25
Tianjin	59	50.85%	2123	26
Shanghai	62	1.61%	2067	16
Guangzhou	30	16.67%	899	22
Shenzhen	25	44.00%	785	26
National mean	43.16	37.82%	511	25

Notes: The statistics are calculated using the large-scale thermal power plants located outside a given city and within 500km. Following Freeman et al. (2019), we define the upwind region as a section of a circular buffer drawn at a distance of 500km from a given city, and the angle between the section and the annual dominant wind direction of the city being at least 45 degrees.

Table A10: Baseline Economy and the Wind Direction IV

Dependent variable:	The ratio of upwind plants		Wind direction and coal plants IV	
	(1)	(2)	(3)	(4)
Baseline Population	-0.0106	-0.0132	0.412	0.226
	(0.0309)	(0.0279)	(0.489)	(0.440)
Baseline GDP per capita	-0.0151	-0.0231	-0.132	-0.414
	(0.0302)	(0.0289)	(0.526)	(0.522)
Baseline Industrial Elec cons	0.0225		0.210	
	(0.0224)		(0.347)	
Baseline Total Elec cons		0.0255		0.414
		(0.0178)		(0.269)
Observations	281	281	281	281
City Controls	Yes	Yes	Yes	Yes

Notes: City level regressions for 281 cities. We drop cities with missing values in baseline characteristics. Dependent variables are based on plants built post 2005, independent variables are measured in the year 2004. Standard errors clustered at the hukou city level are reported in parentheses. City controls include distance to Shanghai, Tianjin and Shenzhen seaports.

Table A11: Placebo Wind Directions

	Log (PM2.5)		Leave hukou city indicator			
			Coal IV Placebo (+90 degrees)		Coal IV Placebo (+180 degrees)	
			Low edu	High edu	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)
Log (PM2.5)			-0.0708	0.0235	-0.159	-0.0281
			(0.0594)	(0.0518)	(0.157)	(0.0760)
Coal IV Placebo	0.00957**					
(wind direction+90 degrees)	(0.00478)					
Coal IV Placebo		0.00134				
(wind direction+180 degrees)		(0.00173)				
Observations	332	332	643,124	118,424	643,124	118,424
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	N	N	Yes	Yes	Yes	Yes
F-test of IVs			6.324	8.940	0.935	1.521

Notes: Individual level regressions across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. 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.

2015; Jans et al., 2018), and as such, been scrutinized thoroughly. Nonetheless, we examine whether lagged pollution levels can predict future levels of the strength of thermal inversions in Table A12. We fail to find any such meaningful associations. Furthermore, we also find that lagged inversions do not predict future inversions, suggesting that levels of auto-correlation in inversions are low, and we may consider the data generating process underlying inversions to be close to random.

A.1.3 The Huai River Regression Discontinuity

Between 1950-1980 China established coal-based free heating systems 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 releasing a large amount of pollutants. Chen et al. (2013) examine the effects of this policy on life expectancy using an RD where they compare cities just north of the river to those just south of it. We find that high-skilled workers are less likely to in-migrate into cities with more pollution (but find no significant impacts on out-migration).

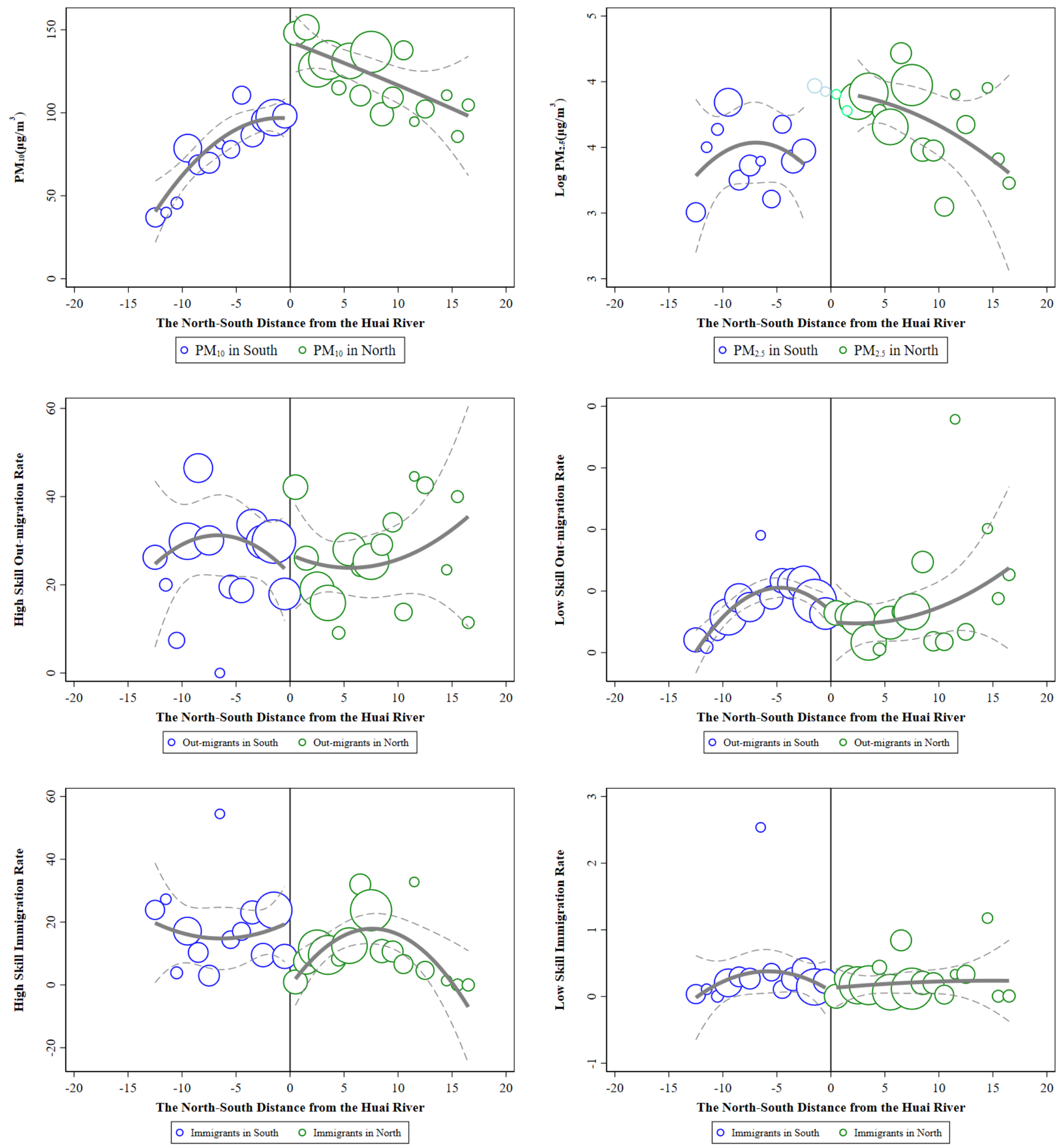
Table A12: Lagged Pollution and Thermal Inversions

	Dependent variable: Strength of inversions		
	(1)	(2)	(3)
Lagged Log(PM2.5)	-11.68	-11.79	-11.38
	(12.88)	(12.77)	(12.80)
Lagged number of inversions			0.241
			(0.181)
Lagged strength of inversions		0.0332	-0.0373
		(0.0330)	(0.0774)
Observations	3,652	3,652	3,652
R-squared	0.961	0.961	0.961
City fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes

Notes: Standard errors clustered at city level are reported in parentheses. City level regressions for 332 cities over 11 years (2005 to 2015). Specifications include city and year fixed effects.

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 PM10 and PM2.5 levels. 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 effect 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: The top row shows the discontinuity in PM10 and PM2.5 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

Here we examine different model specifications, sample restrictions and control variables. First, we employ an individual-level longitudinal panel data, allowing us to track individuals over time and control for individual-level unobservables. Importantly, we use an alternative definition for migration status regardless of *hukou* location.

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. Finally, we use alternative samples, alternative measures of air quality, and alternate sets of co-variate controls to examine the effects of pollution.

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 the China Labor-force Dynamic Survey (CLDS), which is a national social survey targeted at labor force dynamics in China. CLDS 2016 asks a retrospective history of locations for individuals, and we create an individual-level longitudinal panel between 2008 and 2016. 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*. 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.

Table A13 presents the IV estimation of the relationship between pollution and the out-migration tendency of individuals, controlling for year- and individual- fixed effects. Including individual-fixed effects allow us to account for individual-level unobservables (such as taste for clean air, individual preferences for a specific city) that may be correlated with migration decisions. A 10% increase in PM2.5 raises out-migration rates by 2.19 percentage points for high-skilled workers.

In our main table, migrants are defined as those who are away from their *hukou* city, so 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, as migrants with high education attainment find it easier to obtain local *hukou* than those without. We include both non-*hukou* migration (change residential locations without changing *hukou*

Table A13: Individual Longitudinal Panel with Individual Fixed Effects

	Dependent variable: Leave city location indicator								
	Full sample	Low edu	High edu	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	0.115**	0.0812	0.219**	0.0960*	0.0644	0.193**	0.0774	0.0385	0.206**
	(0.0558)	(0.0542)	(0.100)	(0.0540)	(0.0525)	(0.0964)	(0.0539)	(0.0538)	(0.0972)
Observations	122,841	104,184	18,657	122,841	104,184	18,657	122,841	104,184	18,657
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	No	No	No	Yes	Yes	Yes	No	No	No
Regional trend	No	No	No	No	No	No	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	51.62	43.38	55.60	53.91	43.99	59.93	50.67	42.48	56.85

Notes: Individual-level regressions across 277 cities and between 2008-16. The CLDS asks retrospective histories of residential locations, covering 277 cities.Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification using the number of thermal inversions. Demographics include age, age-squared. Weather controls include temperature, wind speed, sunshine duration and humidity. Region trend is a region-specific time trend.

locations) and *hukou* migration (change both residential and *hukou* location) in this definition of migration status. The effects on high-skill migration shown in Table A13 are, as expected, larger in magnitude than our baseline estimates.

We control for weather amenities in the next three columns. The coefficient estimates are similar. To account for the potential role played by differential migration patterns between coastal and inland China, we further add region-specific trends in the last three columns. Including region-specific trends do not affect our results.

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 PM2.5 concentration over different time intervals. Since the coal-fired power plants are essentially leveraging the cross-sectional nature of the data, we use the occurrence of thermal inversions averaged over different time periods to deal with the endogeneity of cumulative air pollution.

In Table A14, we use specifications where PM2.5 are averaged over 5, 15 and 18 years, respectively, leading up to the migration decision. We find that the longer the time period of

Table A14: PM2.5 Measured over Different Time Intervals

	Dependent variable: Leave hukou city indicator								
	Full sample	Low edu	High edu	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (Mean PM2.5: 1998-2015)	0.113***	0.104***	0.173***						
	(0.0236)	(0.0235)	(0.0394)						
Log (Mean PM2.5: 2001-2015)				0.107***	0.0979***	0.168***			
				(0.0230)	(0.0231)	(0.0387)			
Log (Mean PM2.5: 2011-2015)							0.0913***	0.0850***	0.129***
							(0.0220)	(0.0231)	(0.0297)
Observations	761,548	643,124	118,424	761,548	643,124	118,424	761,548	643,124	118,424
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	65.36	63.86	44.57	68.08	66.49	45.59	91.53	87.48	71.38

Notes: Individual level regressions across 332 cities.Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables using the number of thermal inversions averaged over different time intervals. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. 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.

PM2.5 exposure, the larger the response in small increments. As such, pollution exposure in a short time frame has a slightly smaller impact than the same amount of exposure spread out over a longer time period. Once again, we see similar empirical patterns as our baseline estimates. The effects of cumulative pollution on emigration are also more pronounced for the skilled than the unskilled.

A.2.3 Different Samples and Alternative Measure of Air Quality

We use different samples to explore the relationship between pollution and out-migration. We then use an alternative measure of air quality as our independent variable. First, we examine whether big cities, high polluting cities, or province capitals are driving our results. We exclude such cities one at a time in Panel A and columns 1-6 in Panel B of Table A15. This may help allay concerns about the influence of major cities or capitals in pollution policy, placement of plants, or being outliers in terms of pollutants and/or skilled jobs.

Next, we may expect that coal-fired plants plants are more likely to be located in coal producing areas, affecting the underlying industrial structure, and raising concerns about other unobservable associations with migration rates. Shanxi is the largest coal producing province in China, which we exclude in our estimation. As reported in columns 7-9 in Panel B of Table A15, the results are slightly more precisely estimated.

Table A15: Without Big Cities, High Polluters and Coal Producing Regions

Panel A:	Exclude Beijing			Exclude Tianjin			Exclude Shijiazhuang		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	0.0761*	0.0609	0.132***	0.142**	0.129**	0.180***	0.0783**	0.0623	0.140***
	(0.0389)	(0.0424)	(0.0369)	(0.0556)	(0.0619)	(0.0487)	(0.0393)	(0.0426)	(0.0390)
Observations	752,993	638,529	114,464	745,903	632,324	113,579	758,412	640,631	117,781
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	52.58	46.35	42.31	31.46	27.67	29.28	51.78	45.94	40.28

Panel B:	Exclude Shenyang			Exclude Zhengzhou			Exclude Cities in Shanxi Province		
	Full sample	Low edu	High edu	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	0.0766**	0.0605	0.138***	0.0816**	0.0641	0.155***	0.0921***	0.0773**	0.148***
	(0.0387)	(0.0421)	(0.0377)	(0.0397)	(0.0431)	(0.0377)	(0.0349)	(0.0375)	(0.0345)
Observations	758,589	641,096	117,493	754,535	637,980	116,555	722,306	609,605	112,701
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	53.42	47.25	42.65	51.02	45.21	40.14	84.82	77.61	68.94

Notes: Individual level regressions across 331 cities in Panel A and column 1-6 in Panel B (we exclude one big city at a time) and across 321 cities in column 7-9 in Panel B (we drop cities in Shanxi). Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and urban hukou indicator.

In Table A16, we split up the sample by age groups and marital status. We see that the implications of pollution on emigration are stronger for younger workers with higher education. Panel B further shows that married individuals are more responsive, perhaps as they may have young children that may be affected by poor air quality.

Finally, we turn our attention to an alternative measure of air quality. As sources of pollution affect not just PM2.5 but also other pollutants, we may pick up the combined impact of many pollutants. 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. Furthermore, the AQI is officially reported and widely disseminated. Table A17 shows similar results using the AQI as our independent variable of interest.

Table A16: By Age Groups

	Dependent variable: Leave hukou city indicator							
	Age<35		Age>=35		Age<35&Married		Age<35&Unmarried	
	Low edu	High edu	Low edu	High edu	Low edu	High edu	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (PM2.5)	0.0640	0.193***	0.0572	0.0684***	0.0922	0.253***	0.00214	0.140***
	(0.0627)	(0.0504)	(0.0368)	(0.0227)	(0.0737)	(0.0713)	(0.0435)	(0.0370)
Observations	199,958	70,643	443,166	47,781	66,956	37,053	133,002	33,590
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	33.65	40.70	49.93	40.62	34.24	36.70	30.24	42.93

Notes: Individual level regressions across 332 cities.Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A17: Air Quality Index

	Dependent variable: Leave hukou city indicator		
	Full sample	Low edu	High edu
	(1)	(2)	(3)
Log (AQI)	0.0795	0.0581	0.176**
	(0.0659)	(0.0682)	(0.0736)
Observations	708,482	595,958	112,524
City Controls	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
F-test of IVs	36.16	34.99	29.59

Notes: Individual level regressions across 238 cities that report AQI in 2014. Independent variable is Log (Annual mean Air Quality Index).Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. 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.

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 Table A18 we examine robustness to additional geological and meteorological controls when using the thermal inversions instruments. These include measures of rainfall, average slope (elevation) of the city, fraction of the region covered by water bodies, and whether the city is coastal or not.

In Table A20 we test the validity of the wind IV to other economic and industrial controls. In the first few columns of Table A20 Panel A, we account 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.

Table A18: Additional Weather and Geological Controls: For Thermal Inversions

	No. of Inversions			Strength of Inversions		
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)
Log(PM 2.5)	0.117***	0.115**	0.155***	0.0880*	0.0842	0.151**
	(0.0443)	(0.0469)	(0.0483)	(0.0514)	(0.0541)	(0.0599)
Observations	685,699	575,693	110,006	685,699	575,693	110,006
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
F-Test of ivs	37.10	34.22	35.23	13.80	13.06	13.13

Notes: Individual level regressions across 261 cities (for which we have measures of the controls). Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specification using thermal inversions. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Weather controls include temperature, wind speed, sunshine duration and humidity. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. In addition, we control for annual rainfall, average slope (elevation) of the city, fraction of the region area under water coverage, and whether on the coast.

In the next three columns of Table A20 Panel A, we include controls for local economic production in 2000. Our economic controls include baseline measures of GDP per capita, as well as the ratio of product values of services and manufacturing as proxies for the industrial structure. 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 three columns of Table A20 Panel A. The inclusion of these industrial emissions does little to affect the impacts of PM2.5 concentration.

In the first three columns of Panel B, we start with all three sets of controls that we have in Panel A entered simultaneously. After which, we address the concern that power

Table A19: Additional Weather and Geological Controls: For House Price Regressions

	Log(House Prices) (1)	Log(House Prices) (2)	Log(House Prices) (3)
Log(Population)$_d$	0.259**	0.314**	0.204*
	(0.121)	(0.147)	(0.119)
Log(L_{hd}/L_{ud})	0.423***	0.407***	0.456***
	(0.0583)	(0.0690)	(0.0556)
Observations	121	121	117
Weather Control	No	Yes	Yes
Geological Control	No	No	Yes
F-Test of ivs	12.66	13.14	11.73

Notes: City-level regressions, and robust standard errors. We report Kleibergen-Paap rk Wald F statistic. Instrumental variables specifications, where we instrument for log population and log skill ratio as described in the text and Table 5. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Weather controls include temperature, wind speed, sunshine duration and humidity. Geographic controls include average slope (elevation) of the city, fraction of the region area under water coverage, and whether on the coast.

plants may be built near cities that require more electricity. Even though those plants supply electricity to vast areas including many remote provinces (Freeman et al., 2019), we examine this concern by controlling for city-level electricity consumption. We include total (industrial, commercial and residential) electricity consumption in the next three columns of Table A20 Panel B, and add industrial electricity consumption in the last three columns. Adding city-level electricity consumption barely changes our empirical patterns.

B Air Pollution Data Disclosure in China

Despite hazardous levels of pollution, Chinese citizens used to have limited or no access to information about local air quality. In 2008 the US embassy published PM2.5 data from five Chinese cities, leading to a public attention on air quality. In response to public demand for the publication of PM2.5 data, the Chinese government started to disclose real time PM2.5 data in most Chinese cities from 2012. Information on real time PM2.5 was made available in all Chinese cities by January 1, 2015.

The disclosure of pollution information had an important effect on household avoidance behavior. The sales of indoor air filtration more than doubled in response to the PM2.5 data disclosure in 2012. Table B1 shows that PM2.5 data disclosure had significant impacts on the out-migration elasticity of the high-skilled.

Table A20: Additional Controls: Skills, Economic and Industrial Emissions

Panel A	Dependent variable: Leave hukou city indicator								
Additional Controls:	Baseline skill controls			Baseline economy controls			Industrial emmisions controls		
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	0.0712	0.0592	0.124***	0.0403	0.0251	0.0984**	0.0539	0.0396	0.119***
	(0.0433)	(0.0466)	(0.0427)	(0.0422)	(0.0454)	(0.0444)	(0.0435)	(0.0470)	(0.0373)
Observations	674,032	565,239	108,793	674,032	565,239	108,793	674,032	565,239	108,793
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	55.68	53.73	33.41	51.43	48.91	33.90	60.76	59.45	36.89

Panel B	Dependent variable: Leave hukou city indicator								
Additional Controls:	All 3 sets of controls			All 3 sets + Total elec cons			All 3 + Industry elec cons		
	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu	Full Sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (PM2.5)	0.0756**	0.0677*	0.109***	0.0653*	0.0587	0.0959***	0.0643*	0.0574	0.0972***
	(0.0343)	(0.0362)	(0.0335)	(0.0369)	(0.0393)	(0.0346)	(0.0358)	(0.0380)	(0.0338)
Observations	674,032	565,239	108,793	650,828	544,681	106,147	650,828	544,681	106,147
City Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	58.67	58.10	34.15	53.59	52.68	31.95	54.78	54.32	31.76

Notes: Individual level regressions across 261 cities in Panel A, and columns 1-3 of Panel B (cities for which we have measures of the controls). Individual-level regressions across the 253 cities for which we have electricity consumption data and the full set of city controls, in columns 4-9 of Panel B.Standard errors clustered at the hukou city level are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic. 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 distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. Total electricity consumption includes industrial, residential and commercial consumption.

Table B1: PM2.5 Data Disclosure and Outmigration

Dependent variable:	Leave city location indicator					
	Full sample	Low edu	High edu	Full sample	Low edu	High edu
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Log(PM2.5)}_{ot} \times$ Data disclosed indicator_{ot}	0.000902	-0.000585	0.0106**	0.000997	-0.000482	0.0105**
	(0.00147)	(0.00139)	(0.00495)	(0.00143)	(0.00130)	(0.00488)
Observations	122,841	104,184	18,657	122,841	104,184	18,657
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Regional trend	No	No	No	Yes	Yes	Yes

Notes: Individual-level regressions across 277 cities. The CLDS lists the history of residential locations, which include 277 cities. Standard errors clustered at the city level. We report Kleibergen-Paap rk Wald F statistic. Demographics include age, age-squared. We control for the log of PM2.5 and an indicator for PM2.5 data disclosure. Data disclosure indicator=1 if PM2.5 data was officially made available in a given city-year, =0 otherwise.

C Additional Tables and Figures

Table C1: Examples of the Points-based *Hukou* Policy Across Chinese Cities

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

Note: Table shows a few examples of *hukou* requirements for city workers. Data come from the official government websites of the Beijing, Shanghai, Guangzhou and Shenzen. 'Total *hukou* points needed' show the number of points required to obtain local *hukou*, whereas the lower rows show how different qualifications add to the individual's points.

Table C2: Preferences for Environmental Issues by Education Levels

	Claim environmental issue is terrible (1)	Discuss environmental issues (2)	Donation for environment protection (3)	Concern over environmental issue (4)	Appeal on Environmental issue (5)	Government environmental activity (6)	Non-government environmental activity (7)
High school	0.130***	0.135***	0.0592***	0.121***	0.0252***	0.121***	0.0691***
	(0.0116)	(0.0173)	(0.0115)	(0.0153)	(0.00841)	(0.0138)	(0.0116)
Some college or above	0.173***	0.233***	0.152***	0.176***	0.0571***	0.248***	0.157***
	(0.0118)	(0.0190)	(0.0154)	(0.0191)	(0.0120)	(0.0168)	(0.0143)
P-value	0.001	0.000	0.000	0.000	0.010	0.000	0.000
t-value	-3.20	-6.10	-5.65	-3.58	-2.62	-8.16	-6.25
Baseline average	0.497	0.376	0.106	0.389	0.062	0.133	0.100
City FE	Y	Y	Y	Y	Y	Y	Y
Demographics	Y	Y	Y	Y	Y	Y	Y
N	29169	11331	11331	11331	11331	11331	11331
adj. R2	0.085	0.202	0.182	0.202	0.186	0.196	0.203

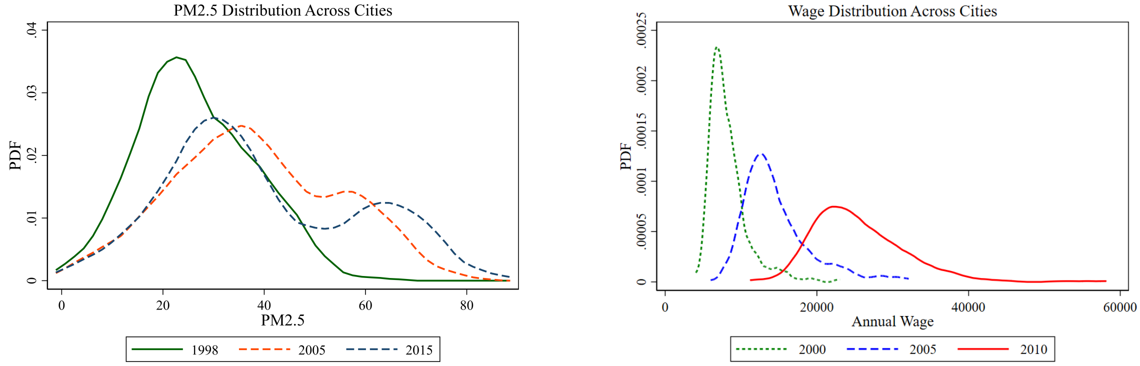
Note: Individual level regressions across 113 cities in column 1 and 127 cities in column 2-7. Demographic controls include age, age-squared, gender, marital status and an indicator for urban *hukou*. The data source is the China Household Panel Survey (CFPS) in column 1 and the Chinese General Social Survey (CGSS) in column 2-7. In the CFPS , 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. 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. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table C3: First-stage Relationships for Additional Model-based Parameter Estimation

Panel A: For Labor Supply	Log (Unskilled Real Wage)		Log (Skilled Real Wage)	
	(1)	(2)	(3)	(4)
NTR IV	0.161***	0.162***		
	(0.0161)	(0.0153)		
WID IV			0.00222***	0.00387***
			(0.000123)	(0.000111)
Pollution IV	No. of Inversions	Strength of Inversions	No. of Inversions	Strength of Inversions
Observations	13,570	13,570	13,570	13,570
R-squared	0.132	0.132	0.136	0.153
Panel B:	For Agglomeration		For House Prices & Pollution	
	Log (Number of Skilled Workers)		Log(Skill Ratio)	Log(Population)
	(1)	(2)	(3)	(4)
ΔCollege graduates$_{2001-5}$	0.400***	0.395***		
	(0.0386)	(0.0390)		
ΔCollege graduates$_{2001-5}$/Baseline $L_{u,2001}$			0.835***	0.127
			(0.130)	(0.0915)
Predicted migration flow			0.00158	0.00425***
			(0.00148)	(0.00102)
Pollution IV	No. of Inversions	Strength of Inversions		
Observations	121	121	121	121
R-squared	0.672	0.665	0.573	0.546

Notes: We control for number of inversions in Column 1 and 3, and strength of inversions in Column 2 and 4. All regressions weighted by the population in 2000. Panel A: We control for hukou index, the interaction of hukou index to migration status indicator, the inverse hyperbolic sine of distance from origin to destination cities. We also control for temperature, humidity, sunshine duration, and wind speed. Standard errors clustered at the city level are reported in parentheses. Panel B: We control for distance to seaport, region fixed effects, city area and weather amenities (temperature, humidity, sunshine duration and wind speed). Robust standard errors are reported in parentheses. We report Kleibergen-Paap rk Wald F statistic.

Figure C1: The Distribution in Pollution and Wages Across Cities

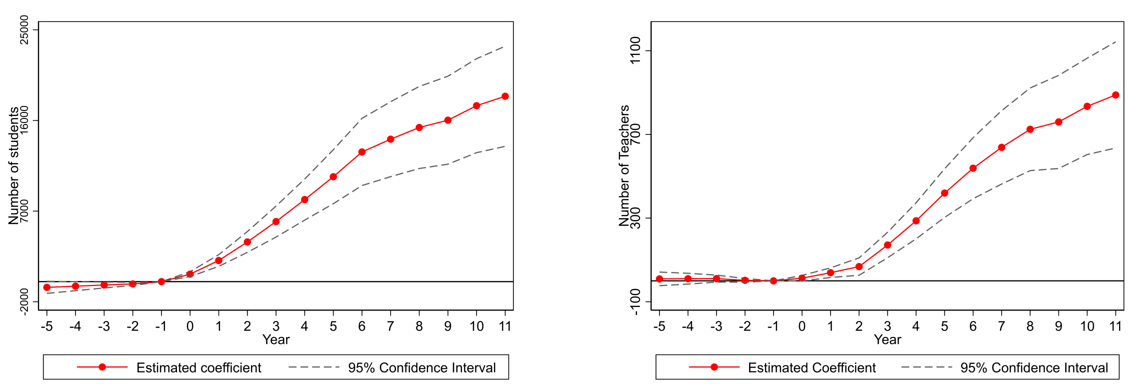


(a) Spread in Pollution Across Cities (b) Wage Distribution Across Cities

Notes: Distributions across cities for different years. Wage distribution drawn from the City Statistical Yearbooks. PM2.5 data from the Global Annual PM2.5 grids.

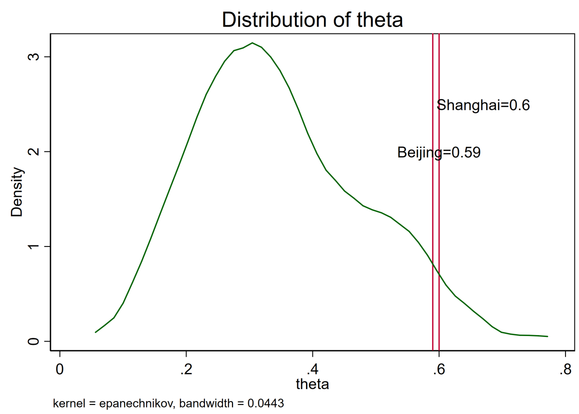
Figure C2: Event Study of the Number of College Students by Baseline Propensity to Expand

(a) College Students in City (b) University Professors in City



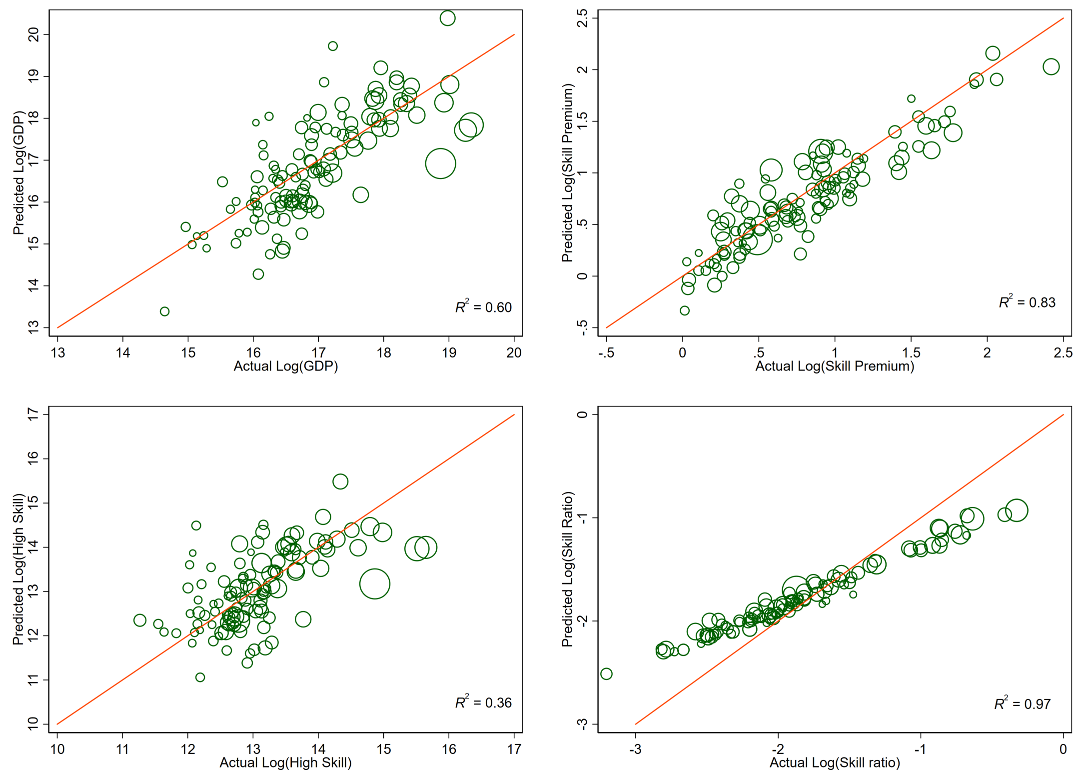
Notes: We test for pre-trends and dynamics of the college expansion policy in an event study framework. City-by-year level regressions from 1994 to 2010. Outcome is number of new college students, relative to one year preceding the expansion (1998). In Figure C2a, we run the regression: $Students_{dt} = \beta_0 + \sum_{t=1994}^{2010} \beta_t \left(1_{t-1998} \times Treat_{d,1990}\right) + \gamma_t + \delta_c + \varepsilon_{dt}$. In Figure C2b the dependent variable is the number of professors in the city. We plot β_t. $Treat_{d,1990}$ is the number of university students in 1990. Horizontal axis is normalized to the year preceding the expansion (1998).

Figure C3: Distribution of θ_{hd} across cities, from Equation 18



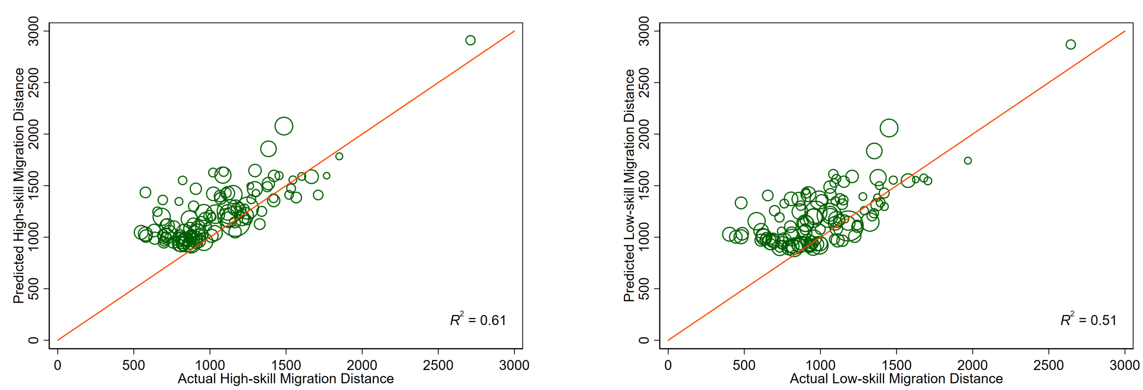
Graph describes the distribution of θ_{hd} across cities. Equation 18 describes how we estimate θ_{hd}.

Figure C4: Model Fit in 2015



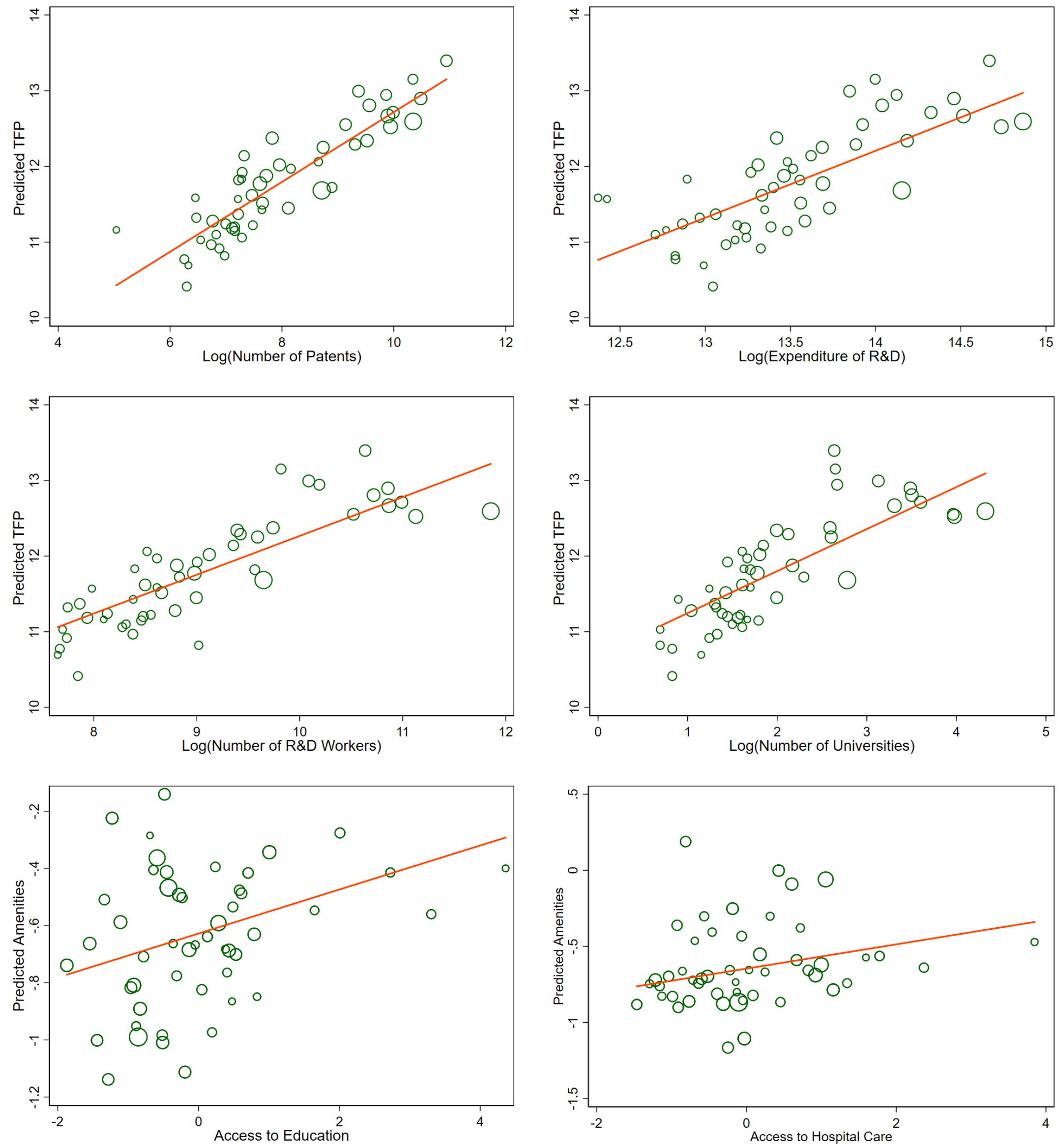
Notes: We plot the actual and predicted relationship between our main variables, where the predictions are based on model-estimated parameters. Bubbles are weighted by populations in the year 2000. We exclude the residual ξ_{sod} in our model predictions.

Figure C5: Out of Sample Test of Migration Distance



Notes: We plot the relationship between predicted and actutal migration distance for migrants who move out from their *hukou* regions, where the predictions are based on model-estimated parameters. Bubbles are weighted by populations in the year 2000. We exclude the residual ξ_{sod} in our model predictions.

Figure C6: Correlates of Predicted TFP and Predicted Amenities



Notes: We plot the relationship between predicted TFP and various innovation measures in the top four figures, and plot the relationship between predicted amenities and access to education and between predicted amenities and access to hopspital care in the bottom two figures. Access to education is measured by the first principal component of the number of schools per capita and the number of teachers per capita. Access to hospital care is measured by the first principal component of the number of hopitals per capita and the number of doctors per capita. Bubbles are weighted by populations in the year 2000. We exclude the residual ξ_{sod} in our model predictions.

D Additional Counterfactual Results

Table D1: Additional Counterfactual Results: Reducing Pollution in Beijing

	Panel A: GDP per Worker in Nearby Cities (within 300km from Beijing) (%)			
	Overall Change	Health Channel	Relocation	Relocation+Agglom
	(1)	(2)	(3)	(4)
Reduce steady state PM2.5	-0.072	0.000	-0.025	-0.072
Relax skilled hukou	-0.061	0.000	-0.023	-0.061
Relax unskilled hukou	-0.008	0.000	0.005	-0.008
Reduce PM2.5 & relax skilled hukou	-0.086	0.000	-0.021	-0.086
Reduce PM2.5 & relax unskilled hukou	-0.181	0.000	-0.058	-0.181
	Panel B: GDP per Worker in Beijing Under No Externalities or Pure Relocation (%)			
	Overall Change (no externalities)	Health Channel (no externalities)	Relocation (no externalities)	Relocation (pure relocation)
	(1)	(2)	(3)	(4)
Reduce steady state PM2.5	10.364	5.819	4.295	4.111
Relax skilled hukou	4.931	0.000	4.931	4.931
Relax unskilled hukou	-4.654	0.000	-4.654	-4.654
Reduce PM2.5 & relax skilled hukou	15.116	5.819	8.785	8.628
Reduce PM2.5 & relax unskilled hukou	5.707	5.819	-0.106	-0.304

Notes: In panel A, we replicate our exercises in Table 7 and present the effects on nearby cities . Panel A column 1 shows the gain to overall GDP per worker. Panel A column 2 shows the component purely explained by the health-productivity channel. Panel A column 3 shows the pure relocation channel, and Panel A column 4 also incorporates agglomeration as a consequence of relocation. In panel B columns 1-3, we do addiontal excercises, shutting down agglomeration and congestion effects. Panel B Column 1 shows the gain to overall GDP per worker. Panel B column 2 shows the component purely explained by the health-productivity channel. Panel B column 3 shows the pure relocation channel. In panel B column 4 we further shut down how the health effects of pollution may affect relocation choices. We reduce the steady state amount of pollution in Beijing by 50% (row 1 in panels A and B). Next, we lower *hukou* restrictions for each skill level (rows 2 and 3 in panels A and B) by 50%. Finally, we lower the *hukou* regulations by 50% while reducing steady state pollution (rows 4 and 5 in panels A and B).

Table D2: Distributional Consequences of Reducing Pollution in One City

	Skilled Wage in Beijing			Unskilled Wage in Beijing		
	Overall	Health	Relocate+Agglom	Overall	Health	Relocate+Agglom
	(1)	(2)	(3)	(4)	(5)	(6)
Reduce steady state PM2.5	2.702	5.819	-2.945	18.260	5.819	11.757
Reduce exogenous part of PM2.5	2.589	5.344	-2.616	16.699	5.344	10.779
Relax skilled hukou	-4.332	0.000	-4.332	12.462	0.000	12.462
Relax unskilled hukou	7.661	0.000	7.661	-6.371	0.000	-6.371
Reduce PM2.5 & relax skilled hukou	-1.580	5.819	-6.992	33.108	5.819	25.789
Reduce PM2.5 & relax unskilled hukou	10.438	5.819	4.365	10.559	5.819	4.479

Notes: This table shows the corresponding wage effects of the policy exercises in our main Table 7. 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 lower the *hukou* restrictions for each skill level (rows 3 and 4) by 50%, keeping pollution fixed. Finally (rows 5 and 6) we relax *hukou* regulations by 50% 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.

Table D3: The Productivity Effect of Relocating Pollution (no externalities and pure relocation)

	Change in GDP per Worker (%)			
	Overall changes (no externalities)	Health (no externalities)	Relocation (no externalities)	Relocation (pure relocation)
	(1)	(2)	(3)	(4)
Relocate steady state PM2.5	7.559	2.893	3.097	2.332
Relax hukou	2.318	0.000	2.318	2.318
Relax overall mobility constraints	6.319	0.000	6.319	6.319
Relocate PM2.5 & relax hukou	9.845	2.893	5.216	4.486
Relocate PM2.5 & lower migration costs	14.143	2.893	9.105	8.418

Notes: In this table, we do an additional exercise, shutting down agglomeration and congestion effects. 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. In column 4 we further shut down how the health effects of pollution may affect relocation choices. We relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relax the *hukou* restriction in the 24 top tier cities by 50% (row 2). In row 3 we relax overall migration costs in the 24 high tier cities by 50%.

Table D4: Distributional Effects of Relocating Pollution Across Cities

	Skilled Wage			Unskilled Wage		
	Overall	Health	Relocate+Agglom	Overall	Health	Relocate+Agglom
	(1)	(2)	(3)	(4)	(5)	(6)
Relocate steady state PM2.5	17.723	4.493	10.484	1.070	1.638	-1.917
Relocate exogenous part of PM2.5	8.924	2.402	5.688	0.983	1.014	-0.469
Relax hukou	3.861	0.000	3.861	1.933	0.000	1.933
Relax overall mobility constraints	10.066	0.000	10.066	5.384	0.000	5.384
Relocate PM2.5 & relax hukou	19.923	4.493	12.649	2.403	1.638	-0.656
Relocate PM2.5 & lower mig costs	27.784	4.493	19.864	6.701	1.638	3.166

Notes: This table shows the corresponding wage effects of the policy exercises in our main Table 8. 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 24 top tier cities by 50% (row 3). In row 4 we relax overall migration costs in the 24 high tier cities by 50%. Columns 1-3 show the effects on skilled workers, while columns 4-6 show the effects on unskilled workers.

Table D5: Welfare Effects of Relocating Pollution Across Cities

	Skilled Welfare	Unskilled Welfare	Average Welfare
	(1)	(2)	(3)
Relocate steady state PM2.5	29.660	2.258	4.628
Relocate exogenous part of PM2.5	13.122	1.326	2.306
Relax hukou	9.895	0.935	1.710
Relax overall mobility constraints	17.458	10.051	10.692
Relocate PM2.5 & relax hukou	42.911	1.502	5.085
Relocate PM2.5 & lower migration costs	55.354	13.289	16.928

Notes: This table shows the corresponding welfare effects of the policy exercises in our main Table 8. 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, we also relax *hukou* restrictions in the 24 top tier cities by 50% (row 3). In row 4 we relax overall migration costs in the 24 high tier cities by 50%. Rows 3 and 4 keep pollution fixed. Rows 5 and 6 relocate pollution while relaxing migration costs.

Table D6: Distributional Effects of Pollution Regulation

	Skilled Wage			Unskilled Wage		
	Overall	Health	Relocate+Agglom	Overall	Health	Relocate+Agglom
	(1)	(2)	(3)	(4)	(5)	(6)
Control PM2.5	3.504	1.485	1.896	3.603	1.536	1.958
Control PM2.5 & relax hukou	7.583	1.485	5.840	5.645	1.536	3.921
Control PM2.5 & lower mig costs	13.964	1.485	12.115	9.142	1.536	7.368

Notes: This table shows the corresponding wage effects of the policy exercises in our main Table 9. 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 (row 2) and migration costs (row 3) in higher tier cities by 50%. Columns 1-3 show effects on skilled wages. Columns 4-6 on unskilled wages.

Table D7: Welfare Effects of Pollution Regulation

	Skilled Welfare	Unskilled Welfare	Average Welfare
	(1)	(2)	(3)
Control PM2.5	7.150	1.087	1.611
Control PM2.5 & relax hukou	18.425	1.978	3.401
Control PM2.5 & lower migration costs	26.668	11.224	12.560

Notes: This table shows the corresponding welfare effects of the policy exercises in our main Table 9. We reduce pollution according to the 2013 *Air Pollution Prevention and Control Plan* (row 1). In addition to pollution regulations, we also relax the hukou restriction (row 2) and migration costs (row 3) in higher tier cities by 50%.

Table D8: Reducing Pollution in Beijing: Changes in GDP per worker (%)

	Panel A: Use Tombe and Zhu (2019) elasticity for all workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Reduce steady state PM2.5	13.984	13.648	13.528	13.422
Reduce exogenous part of PM2.5	12.387	12.005	11.86	11.724
Relax skilled hukou	7.444	7.971	8.194	8.419
Relax unskilled hukou	-4.376	-4.384	-4.359	-4.315
Reduce PM2.5 & relax skilled hukou	21.825	22.07	22.201	22.353
Reduce PM2.5 & relax unskilled hukou	9.516	9.173	9.077	9.012
	Panel B: Use Tombe and Zhu (2019) elasticity for skilled workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Reduce steady state PM2.5	14.37	14.374	14.381	14.392
Reduce exogenous part of PM2.5	12.784	12.767	12.765	12.765
Relax skilled hukou	7.718	8.552	8.907	9.268
Relax unskilled hukou	-3.962	-3.617	-3.467	-3.309
Reduce PM2.5 & relax skilled hukou	22.482	23.361	23.745	24.14
Reduce PM2.5 & relax unskilled hukou	10.333	10.703	10.869	11.044
	Panel C: Use Tombe and Zhu (2019) elasticity for unskilled workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Reduce steady state PM2.5	14.007	13.604	13.422	13.23
Reduce exogenous part of PM2.5	12.394	11.981	11.798	11.607
Relax skilled hukou	7.901	7.587	7.441	7.284
Relax unskilled hukou	-4.179	-4.553	-4.693	-4.819
Reduce PM2.5 & relax skilled hukou	22.331	21.615	21.287	20.938
Reduce PM2.5 & relax unskilled hukou	9.748	8.951	8.621	8.296

Notes: In this counterfactual 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 lower the hukou restrictions for each skill level (rows 3 and 4) by 50%, keeping pollution fixed. Finally (rows 5 and 6) we relax hukou regulations by 50% while reducing steady state pollution.

Table D9: Relocating Pollution for China: Changes in GDP per worker (%)

	Panel A: Use Tombe and Zhu (2019) elasticity for all workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Relocate steady state PM2.5	6.386	5.03	4.339	3.556
Relocate exogenous part of PM2.5	3.592	2.966	2.644	2.273
Relax hukou	2.789	3.601	4.043	4.572
Relax overall mobility constraints	7.56	9.243	10.102	11.091
Relocate PM2.5 & relax hukou	9.054	8.416	8.124	7.828
Relocate PM2.5 & lower migration costs	14.153	14.277	14.348	14.455
	Panel B: Use Tombe and Zhu (2019) elasticity for skilled workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Relocate steady state PM2.5	6.957	6.49	6.289	6.082
Relocate exogenous part of PM2.5	3.812	3.552	3.44	3.325
Relax hukou	2.449	2.697	2.804	2.913
Relax overall mobility constraints	6.708	7.178	7.375	7.574
Relocate PM2.5 & relax hukou	9.298	9.008	8.885	8.759
Relocate PM2.5 & lower migration costs	13.897	13.772	13.715	13.653
	Panel C: Use Tombe and Zhu (2019) elasticity for unskilled workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Relocate steady state PM2.5	6.106	5.288	4.865	4.38
Relocate exogenous part of PM2.5	3.435	3.115	2.95	2.763
Relax hukou	2.962	3.412	3.627	3.862
Relax overall mobility constraints	7.882	8.905	9.371	9.866
Relocate PM2.5 & relax hukou	8.906	8.521	8.306	8.049
Relocate PM2.5 & lower migration costs	14.115	14.27	14.291	14.279

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 row 3, we relax hukou restrictions in the 24 top tier cities by 50%, keeping pollution fixed. In row 4 we relax overall migration costs to the 24 top tier cities by 50%, keeping pollution fixed.

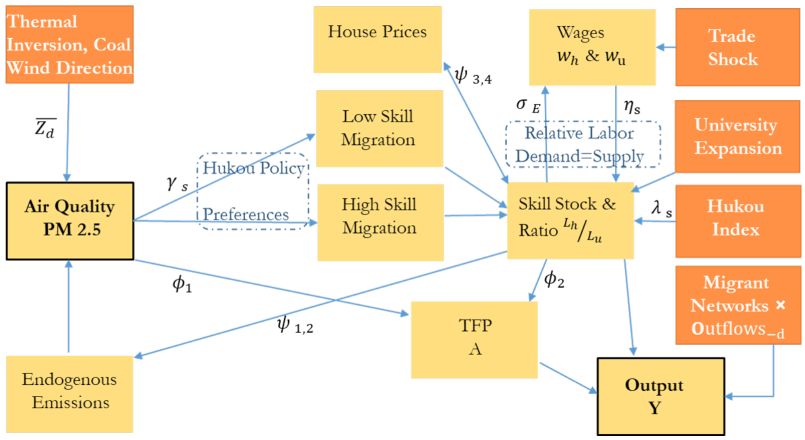
Table D10: 2013 City-level Pollution Caps: Changes in GDP per Worker (%)

	Panel A: Use Tombe and Zhu (2019) elasticity for all workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Reduce steady state PM2.5	3.892	4.282	4.469	4.673
Reduce exogenous part of PM2.5	6.853	8.104	8.756	9.515
Reduce PM2.5 & relax unskilled hukou	11.722	13.864	14.944	16.174
	Panel B: Use Tombe and Zhu (2019) elasticity for skilled workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Reduce steady state PM2.5	3.571	3.567	3.565	3.561
Reduce exogenous part of PM2.5	6.163	6.414	6.52	6.628
Reduce PM2.5 & relax unskilled hukou	10.509	10.987	11.184	11.384
	Panel C: Use Tombe and Zhu (2019) elasticity for unskilled workers			
	Elasticity=1.19	Elasticity=1.4	Elasticity=1.5	Elasticity=1.61
Reduce steady state PM2.5	3.895	4.275	4.454	4.647
Reduce PM2.5 & relax skilled hukou	7.035	7.902	8.311	8.755
Reduce PM2.5 & relax unskilled hukou	12.057	13.508	14.172	14.879

Notes: In this counterfactual 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 high-tier cities (row 2), and overall migration costs (row 3) by 50

E Additional Model Derivations

Figure E1: The Flowchart of Our Spatial GE Model



Notes: We summarize our model's primary relationships in this flowchart for convenience.

E.1 Deriving Labor Supply and Welfare

In this appendix we derive the labor supply curve from the worker utility function in our main text.

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

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

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

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

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

So the derivative of the CDF is given by:

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

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

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

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 (in our main text):

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

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

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

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:

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

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 migration costs to 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 main text, by using the above set of equations:

$$log\, \pi_{sod} = -\eta_s log \overline{V_{so}} + \eta_s \left(log\, w_{sd} - \nu_s log\, hp_d\right) + \eta_s log\, a_{sd} - \eta_s \gamma_s log Z_d - \eta_s M_{sod} + \widetilde{\xi_{sod}} , \tag{10}$$

E.2 Elasticity of Capital, and 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 skill-premia. The average earnings for a worker with skill s in district d would be:

$$\log w_{sd} = \log\left(\frac{\partial Y_d}{\partial \ell_{sd}}\right) = \log\theta_{sd} + \log\rho + \left[1 - \left(1 - \frac{1}{\sigma_E}\right)\left(\frac{1}{\rho}\right)\right]\log Y_d$$
$$+ \left(1 - \frac{1}{\sigma_E}\right)\left(\frac{1}{\rho}\right)(logA_d + (1-\rho)\log\bar{K}_d) - \frac{1}{\sigma_E}\log L_{sd} \quad \text{(E.7)}$$

Here the modified term $\left(\frac{1}{\rho}\right)(logA_d + (1-\rho)\log\bar{K}_d)$ is common across skill levels, and not affect skill premia. It varies across cities, just as TFP in the main model. We can similarly (re)define a modified TFP term that includes the immobile capital.

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

$$Y_d = A_d L_d^{\rho} K_d^{(1-\rho)} \quad \text{(E.8)}$$

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

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

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 E.11, instead of equation 4 in our main text:

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

E.3 On Existence and Uniqueness of the Equilibrium

Section 5 presents the model and equilibrium. Here we describe the determinants behind the existence and uniqueness of the equilibrium defined in Section 5.3.

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 quasi-symmetric. 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 Shanghai 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 Shanghai (say, its large airport), and a component that is Beijing-Shanghai 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.

Additionally, our model contains congestion forces (such as pollution and house prices) and agglomeration (effects on TFP). The existence of multiple equilibria often depends on the relative strength of agglomeration and congestion forces (Allen et al., 2020). An equilibrium is unique if congestion forces are at least as large as the agglomeration forces. That is, the parameters ψ_1, ψ_2, ψ_3 and ψ_4 that determine congestion are meaningful in magnitude, relative to ϕ_1 and ϕ_2 that drive agglomeration. More skilled workers raise TFP (via ϕ_2), yet may lead to more congestion, via 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.

When solving for equilibrium, our model converges to the same unique equilibrium across different starting values. This does not necessarily imply the equilibrium is globally unique. Yet, like other work (Ahlfeldt et al., 2015) we envision that if there were to be multiple equilibria for a different set of parameter values, we would select the counterfactual equilibrium closest to the observed real-world outcome.

F Data Appendix

Pollution Data: In our baseline analysis, 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). We employ PM2.5 satellite data , since it has the following advantages. First, official PM2.5 data were not available for most Chinese cities prior to 2012, and only a subset of cities had access to official PM10 data prior to 2012. However, our satellite PM2.5 data have been available since 1998 for all the cities in China. Second, fine particles (diameter$<$2.5μm) are more hazardous than larger particles (2.5μm$<$diameter$<$10μm) in terms of mortality, cardiovascular and respiratory endpoints, and PM2.5 is considered to be the best indicator of the level of health risks from air pollution. [42] Finally, a potential concern of the official air quality data is that it may be manipulated by the local government (Ghanem and Zhang, 2014; Greenstone et al., 2020), however our satellite data are immune to any underlying data manipulation.

Van Donkelaar et al. (2016) estimate ground-level PM2.5 by combining Aerosol Optical Depth (AOD) retrievals from the NASA MODIS, MISR, and SeaWiFS, which are subsequently calibrated to global ground-based observations of PM2.5 using Geographically Weighted Regression (GWR). The raster grids of this ground calibrated PM2.5 data have a high grid cell resolution of 0.01 degree. This yields a comprehensive and reliable measurement of air quality for a wide range of cities in China, covering all the prefecture, sub-provincial and provincial cities. The correlation between satellite PM2.5 data and monitor-based PM2.5 data in China is up to 0.8 (Freeman et al., 2019). Moreover, our satellite-derived PM2.5 data cover periods before and after China's disclosure of official PM2.5 data, allowing us to analyze the migration reponse to the disclosure of pollution information.

We use the Air Quality Index (AQI) released by Ministry of Environmental Protection (MEP) for robustness checks. AQI is an overall indicator for air pollution concentration calculated using multiple atmospheric pollutants including SO2, NO2, PM10, PM2.5, O3 and CO.

Migration Data: In our baseline analysis, we obtain data on individual migration decisions from the 2015 Population Census of China. China conducts its population census

[42] see WHO report: http://www.who.int/mediacentre/news/releases/2014/air-quality/en/.

every five years; the 2015 census is the latest census with restricted public access. The census records a wide range of demographic and economic characteristics of individuals, including age, gender, education, employment, hukou location, current residential city, and marriage status. We limit the sample to the working age population. In the Census, migrants are defined as those who are away from their hukou city for more than six months. Hukou status determines citizens' access to state-provided goods (such as schools for children) and services (like marriage registries or passport renewals). Given the strong (forced) attachment to one's hukou city, when a person's location of residence differs, it can be reliably characterized as migration.

In addition, we also construct an individual-level longitudinal panel using the China Labor-force Dynamics Survey (CLDS) and use an alternative defination of internal migration to examine the robustness of our empirical pattern. The CLDS data record individual histories of location changes for a sample of 21,086 individuals in 14,226 households across 29 provinces of China. A probability-proportional-to-size sampling (PPS) procedure based on population size and administrative units is adopted to ensure that the survey is nationally representative. As a result, the distribution of sample size across cities in the CLDS is consistent with the geographic distribution of population in China. The CLDS is a national longitudinal social survey, with detailed information on education, work and migration experience. Since the survey asks retrospective migration histories of each individual, we are able to construct a longitudinal panel of location histories between 2008 and 2016. 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. When we analyze migration choices using the longitudinal panel data, we define migration to be an indicator for whether an individual changed city locations between years, regardless of whether they change their hukou status.

In this paper, we refrain from looking at within-city sorting for a few reasons: First, the wage and local labor market are likely to consist of the larger city as a whole. And our spatial structural model analysis aims to quantify the productivity consequences of moving workers from high productivity to low productivity cities in response to relocating pollution. Second, within-city variation can only explain about 19% of the spatial dispersion of air quality in China.

Trade Data: Information on city exports and imports is derived from the China Customs Database, which covers the universe of Chinese exports and imports, and was harmonized and generously provided by the University of California, Davis, Center for International Data (Feenstra et al., 2018). We utilize information on the quantity and value of exports classified by the Harmonized System for all international transactions from China. The data reports the annual trade information on values, quantities, and partner countries at the HS 8-digit level for all Chinese cities in the period under investigation (i.e., 1997 to 2014). As the industry classifications used in tariffs and the China Customs Database (i.e., HS 6-digit) are different from the one in the Annual Survey of Industrial Production (i.e., Chinese Standard Industrial Classification 4-digit), we correspond them to the International Standard Industrial Classification (ISIC) Revision three at the 4-digit level to construct various trade shock measures in practice. We construct a city-level measure that captures the differential impact of PNTR across Chinese prefecture cities based on their pre-2001 industrial activity. Existing industrial activity is measured by each industry's share of total city exports, *prior* to the conferral of PNTR, calculated from the China Customs Database.

Exports are categorized by the destination country and city of origin. The 4-digit city codes provided in the customs data identify a level of geography more disaggregated than the standard prefecture cities in China. Hence, we aggregate city codes in the customs data up to the prefecture level, based on the reported city name. In the end, the original 479 city codes in the customs data are aggregated to 313 prefecture cities, including four municipalities. We do not include exports categorized as process and assembly or process with imported materials, given the heterogeneous response to such processing trade (Yu, 2015).

The tariff data comes from the Trade Analysis and Information System (TRAINS) database, which is maintained by the United Nations Conference on Trade and Development (UNCTAD). The raw tariff data is withdrawn with the simple average at the level of country-HS 6-digit. For additional balance tests, we use data on contract intensity and export licesnses. Using data from Nunn (2007), we measure industry contract intensity in 1997 as the fraction of intermediate inputs employed by firms requiring relationship-specific investments. We use data from Bai et al. (2017) on the fraction of export revenues within an industry covered under export licenses in 2000.

Data on IVs for Pollution: We obtain information on large-scale power plants, their coal consumption, and plant-level electricity generation from China Electric Power Yearbook and China Energy Statistical Yearbook. As in Freeman et al. (2019), the large-scale thermal power plants are defined as the thermal power plants whose installed-capacities are larger than 1 million KW. We mannually collect information on the establishment year of plants, the angle between their locations and annual prevailing wind direction of each city, and the distance to each city.

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

Other Data: We collect data on baseline city characteristics from the City Statistical Yearbooks. The city characteristic varriables include population, GDP, total electricity consumption, industrial electricity consumption, industrial structure (the product value at service sector / manufacture sector), industrial SO2 emission, industrial waste water emission and industrial dust emission, teacher student ratio, the number of hospitals per capita, and the number of doctors per capita.

We obtain data on weather amenities from China Meteorological Data Service Center. Our weather condition variables include temperature, wind speed, sunshine duration and humidity. We manually calcuate the distance from each city to Shanghai seaport, to Tianjin seaport, and to Shenzhen seaport. Since the 2015 Census does not record individual-level wages, we use the CLDS to calculate city-and-education specific average wage.

We measure city-level housing prices using yearly average data on housing rents from the Xitai Real Estate Big Data depository. The data has been collected since 2005, and covered 337 cities 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 statistical yearbooks, and find a correlation of 0.93.

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. Data on the number of college students and graduates at city level are from the China Regional Statistical Yearbook. We create measures of the number of college graduates by city and year, and use the change in graduates from 2001-05 (cohorts just before and after the university expansion policy) as our instrument for the aggolomeration of skilled workers.