

The Productivity Consequences of Pollution-Induced Migration in China

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

July, 2021

Abstract

This paper studies how migration in response to pollution affects aggregate productivity and welfare in China. We document a robust pattern in which skilled workers emigrate away from polluted cities, more than the unskilled. Their greater sensitivity to air quality holds up in cross-sectional variation across cities, panel variation with individual fixed-effects, and when instrumenting for pollution using distant power-plants upwind of cities, or thermal inversions that trap pollution. Pollution therefore changes the spatial distribution of skilled and unskilled workers, which results in higher returns to skill in cities that the educated migrate away from. We quantify the loss in aggregate productivity due to this re-sorting by estimating a model of demand and supply of skilled and unskilled workers. Counterfactual simulations from the estimated model show that reducing pollution would increase productivity through spatial re-sorting by approximately as much as the direct health benefits of clean air. Physical and institutional restrictions on mobility exacerbate welfare losses. People's dislike of pollution explains a substantial portion of the wage gap between cities.

Keywords: Internal migration, air pollution, spatial productivity gaps

JEL Codes: Q52, R12, J61

^{*}University of California – San Diego, gakhanna@ucsd.edu

[†]Jinan University, liang.wenquan@jnu.edu.cn

[‡]Yale University, NBER, Deakin Univ, and CEPR, ahmed.mobarak@yale.edu

[§]Yale-NUS College, ran.song@yale-nus.edu.sg

We thank seminar participants at UCSD, UW (Seattle), NBER Summer Institute (Energy and Environmental Economics; and Urban Economics), Cowles Foundation Conference, LSE Workshop in Environmental Economics, Stanford SITE Migration Conference, U Richmond, UC Berkeley, NUS, Cities and Development Conference (Johns Hopkins), Ortygia Business School, Notre Dame, Georgetown University, Barcelona UAB, the World Bank, Harvard, CUHK, HKUST (China Summer Institute), Nankai-Yale Conference, 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 or sectors within developing countries (Restuccia and Rogerson, 2013, 2017) create an enduring development puzzle: Why do workers remain in low productivity areas when they could experience wage gains elsewhere (Gollin et al., 2014)? It is important to understand the drivers of worker location choices, as policies that permit spatial reallocation could produce substantial productivity gains (Clemens et al., 2019). The literature proposes a few explanations for the low rates of within-country mobility observed across the world: Migration costs (Bazzi, 2017; Bryan and Morten, 2019), income risk at destinations (Bryan et al., 2014; Lewis, 1954), non-transferable location-specific amenities (Munshi and Rosenzweig, 2016), or urban disamenities (Lagakos et al., 2019). We add pollution to this list.

Pollution can have a large effect on where people choose to live and work (Heblich et al., 2020). Particulate matter pollution exceeded WHO air quality guidelines for 96% of Chinese cities in 2015, and on average was four times higher than the level considered safe. Chinese air pollution reduces citizens' life expectancy and causes elevated rates of heart disease, stroke, and lung cancer (Ebenstein et al., 2015, 2017; He et al., 2020). We document that the sharp increases in pollution in China in recent years were concentrated in a few cities, which increased the cross-city dispersion in pollution, and with it, incentives to migrate between cities.

After establishing a causal effect of pollution on emigration, we quantify the aggregate productivity consequences of this movement. While one branch of the literature argues that workers are efficiently sorted (Young, 2014), we show that asymmetric migration responses of skilled and unskilled workers to pollution can create losses in aggregate productivity. This is because skilled workers choose to leave polluted places where they would be more productive, and the production complementarities between skilled and unskilled workers makes the unskilled less productive when the skilled leave. Migration costs, both physical and from Chinese *hukou* policy, differentially restrict mobility by skill. This exacerbates the productivity and welfare losses from pollution.

Pollution and the skill-composition of the workforce are jointly determined, and both depend on other factors such as industrial growth. The first part of our paper is therefore careful to identify the migration response to *exogenous variation* in pollution. To build confidence that our estimates indeed represent the causal effect of air quality on mobility, we go beyond the existing reduced-form literature by assembling several new datasets and investigating this relationship under multiple independent sources of variation. 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), a regression discontinuity around the Huai river (as in Chen et al., 2013), and a meteorological phenomenon called thermal inversions that traps pollution (as in Arceo et al., 2016; Chen et al., 2017; Hicks et al., 2015). We also model *changes in* worker location as a function of pollution *changes*, following individuals over time, and conditioning on individual fixed effects. Across research designs, we find robust evidence that college-educated workers leave areas with higher levels of pollution, while the less educated are comparatively less responsive.

Even then, the reduced form analysis by itself is incomplete, since all parts of the country are affected either directly or indirectly by the re-location of workers, making finding true ‘control groups’ for comparison elusive. Residents of a city experiencing no change in pollution may still see a change in their incentives to migrate, if in-migration from other newly-polluted places changes the equilibrium wage rate in that city. Indeed, the quantities of workers, equilibrium wages, and pollution levels are jointly determined in spatial equilibrium, making it difficult to interpret key parameters solely from reduced-form relationships.

Quantifying the productivity implications of these mobility responses requires a model. For instance, 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, as in Hsieh and Klenow (2009) and 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 that workers have to be paid to reside in polluted cities. This traces out the labor demand curve, and allows us to estimate the elasticity of substitution between the skilled and unskilled. This estimate shows that skilled and unskilled workers display 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 (PNTR) between the US and China, and the change in world import demand (WID), 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 (Andreoni and Levinson, 2001); and (c) worker location decisions may affect agglomeration (Au and Henderson, 2006), (d) house prices (Bayer et al., 2009; Glaeser, 2014a), or (e) the pollution-intensity of production (Glaeser and Kahn, 2010; He et al., 2020). We introduce additional instruments to estimate these elasticities.² Indeed, as workers move, it may change where production takes place (Imbert et al., 2021) 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

¹Pierce and Schott (2016) use the PNTR import shock and Autor et al. (2013) use the WID import shock to document effects on the United States. 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.

²For instance, to estimate skilled-worker agglomeration effects, we leverage a large-scale expansion of university seats that rapidly expanded college enrollment by 20% in certain Chinese cities. Our estimated elasticities are similar to credible estimates in the literature on the direct effect of pollution on productivity (Adhvaryu et al., 2016; Chang et al., 2019; Kahn and Li, 2019), and worker location on agglomeration (Gaubert, 2018; Moretti, 2004). If we were to discipline our model by calibrating elasticities from the literature (instead of estimating them) we would get similar quantitative results.

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 some light on the puzzle we raised at the outset. Indeed, concerns about pollution keep workers away from cities where they could be more productive. The news media has documented this phenomenon,³ but such productivity losses have not been quantified. This phenomenon is also not limited to China: when 9,000 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., 2020). Indeed, 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 halve the ‘exogenous’ component of pollution in Beijing (say, by relocating upwind coal-fired plants), and in another we halve the amount of total pollution (say, by setting pollution caps for Beijing), which would also be a function of factors that are endogenous to our model. In each scenario, GDP per worker rises by more than 12%, mostly as a consequence of skilled workers moving into Beijing, which also raises unskilled wages.

We 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 where industries with greater demand for skilled work like finance, technology or skilled manufacturing are located, to other cities. 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 (e.g. less stringent *hukou* restrictions) further amplifies the effects.

Pollution-control programs recently introduced by the Chinese government suggests that authorities already recognize such gains. The *12th Five-Year Plan on Air Pollution Prevention and Control in Key Regions* sets targets for ambient concentrations of particulate matter, with more stringent targets for high-productivity, polluted regions like Beijing.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

³See 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/>.

⁴See Lustgarten (2020) and “How Climate Migration will Reshape America,” New York Times, Sep 15, 2020, <https://www.nytimes.com/interactive/2020/09/15/magazine/climate-crisis-migration-america.html>.

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

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 worker health and labor 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 air pollution on an economy's prosperity. It is important to incorporate this into policy analysis, as increased pollution and migration have been two of the defining features of the impressive Chinese growth experience over the last 30 years (Brandt et al., 2008; Tombe and Zhu, 2019; Zheng and Kahn, 2013).

Whether relocating pollution also affects aggregate welfare (beyond productivity effects) depends on the precise underlying reason as to why the high and low-skilled react differently to pollution. Survey data shows that this is partly due to different preferences and environmental awareness of the rich, and partly due to differential costs of migrating. Several Chinese cities have adopted a point-based system that exempts workers with skills or higher education from their *hukou* restrictions (see Appendix Table C2). Without the exemption, the system imposes a 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 polluted cities with low wages even as their skilled counterparts leave. Such migration restrictions take other forms around the world, but they are ubiquitous. For example, state-level entitlement schemes in India discriminate against out-of-state migrants and inhibit mobility (Kone et al., 2018). Migration costs are often disproportionately high for the poor in developing countries due to the poor quality of public transport (Bryan and Morten, 2019). 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 environmental economists previously thought. 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.

Beyond the modeling, our empirical work represented several important advances over the contemporaneous reduced-form literature on the link between pollution and migration (Chen, 2020; Yu et al., 2020). For example, Chen et al. (2017) 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. In contrast, the first part of our paper uses 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 (be-

fore and after pollution information was disseminated) to explore the relationship. We use multiple sources of policy-relevant variation, and a battery of sensitivity tests to build confidence. We additionally document the changes in net populations similar to [Chen et al. \(2017\)](#), but recognize them to be the consequence of re-sorting across all cities in general equilibrium. [Xue et al. \(2020\)](#) 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](#)). Very different data and empirical strategies used by these papers produce results consistent with the mechanism highlighted in our paper. But our model helps us answer a question different from all other work: what are the macroeconomic consequences of pollution on production and economic well-being?

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

To build intuition we build a simple framework in [Section 2](#), describe data sources in [Section 3](#), and delineate geographic and temporal patterns of pollution and migration in [Section 4](#). We discuss identification strategies in [Section 5](#) and estimate the causal effect of pollution on migration in [Section 6](#). The quantitative part consists of the theoretical framework in [Section 7](#), estimating model parameters in [Section 8](#), and conducting counterfactual exercises in [Section 9](#). [Section 10](#) concludes.

2 A Simple Framework

To fix ideas, consider a simple framework with just two identical cities ($d = A, B$), and one type of worker. Workers are homogeneous, paid their marginal product w_d , and experience a $utility_d = w_d Z_d^{-\gamma}$. $Z_d^{-\gamma}$ is the level of pollution that reduces utility by a factor γ . Production in each city depends on multiple inputs, and displays a diminishing marginal product of labor.

When pollution, amenities, and inputs into production (labor, capital, technology) in both cities are identical, and there are no migration frictions, then $utility_A = utility_B$ and $w_A = w_B$. As such, there are no gains from worker relocation or pollution reallocation policy.

When pollution levels in city A (exogenously) increase, utility in city A falls, and workers migrate from A to B until welfare between the two cities are equalized. At the end of this migration, $w_B = w_A \left(\frac{Z_B}{Z_A}\right)^\gamma$. This pollution-induced compensating differential leads to a wedge in the marginal products of labor between the two cities, whereby the wage in city A is higher. If we move pollution from A to B , workers migrate from B to A , and the first worker adds $w_A - w_B = w_A \left(\frac{Z_A - Z_B}{Z_A}\right)$ to total output.

As we move pollution to equalize the levels across cities, more workers migrate, and the overall gain to output can be approximated by a [Harberger \(1954\)](#) triangle to be $\frac{\sigma_D}{2} wedge^2$, where σ_D is the absolute value of the slope of the (linearized) labor demand curve, and $wedge = w_A \left(\frac{Z_A - Z_B}{Z_A}\right)$ is the wedge in wages across the two cities.

Now additionally, suppose it is costly to migrate to city B (or there are institutional migration restrictions, like *hukou* policy), then utility in B may be higher, $0 \leq utility_B - utility_A \leq c$, by at most

the migration cost c . So welfare can be improved by relaxing these costs, or relocating pollution from A to B .

Now suppose there are two types of workers that are complements in production. If high-skill workers are mobile but low-skilled workers are not, then as pollution levels rise in city A , high-skill workers leave. In such a scenario, wages for low-skill workers fall as their complementary counterparts leave. The fall in low-skill wages will depend on the elasticity of complementarity between the low and high skilled σ_E . Utility of the low skilled workers in city A will also reduce as they are stuck in a low-wage, high-pollution city. Allowing low skilled workers to migrate would improve both overall productivity (as they match with their complementary inputs), and welfare (as they leave low wage, high pollution cities).

Similarly, suppose city A has more skill-biased capital or technology, that would have made high-skilled h workers more productive in city A . That is, $w_{hA} > w_{hB}$, all else equal. Then relocating pollution away from city A can actually increase overall output once again, as skilled workers move into city A . The first worker that moves, adds an additional $w_{hA} - w_{hB}$ to total output, but also makes the unskilled in A more productive.

A combination of these forces describe how relocating pollution can increase output, how relaxing migration restrictions can improve welfare, and how each depends on the mobility of complementary inputs. We aim to capture these critical insights in our model, but the more complete model we use for quantification will include other important forces as well.

3 Data and Measurement

3.1 Demographic and Migration Data

We measure internal migration using the 2015 Population Census of China. This is the latest census with restricted public access. Importantly, and 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. The census records demographic and economic characteristics of individuals, including education levels, employment details, *hukou* location, and current residential city. We use the 2015 One-Percent Census sample, and restrict our attention to the working age population.

We define migration in a few different ways. First, in the Census, migrants are defined as those who are away from their *hukou* city for more than six months.⁶ *Hukou* status determines citizens' access to state-provided goods (such as schools for children) and services (like marriage registries or passport renewals).⁷ 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.

⁶This definition is consistent with other recent work on internal-migration in China (Combes et al., 2019; Tombe and Zhu, 2019). Most people's a *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*.

⁷*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.

Second, we construct an individual-level longitudinal panel using the China Labor-force Dynamics Survey (CLDS), which records individual histories of location changes for a sample of 21,086 individuals in 14,226 households across 29 provinces of China.⁸ 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 location histories between 2008 and 2016. We define migration to be an indicator for whether an individual changed city locations between years, regardless of whether they change their *hukou* status. The CLDS allows us to account for individual-specific unobservables, track those who have moved multiple times and those who have moved and returned home.

We supplement the migration data with a measure of the stock of workers by skill level computed using the Census. Migration choices ultimately affect the *net* number of skilled and unskilled workers in each city. We show that the number of workers by skill vary systematically with air quality across cities. 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. City-level relationships are weighted by the population in 2000 as it is likely to precede most migration decisions we measure by 2015.

3.2 Air Quality Measures

We use satellite data to measure air quality, which has a few advantages over official sources of pollution data. First, satellite-based PM_{2.5} measures are available for all cities between 1998 and 2015, whereas official PM_{2.5} data are only available since 2012.⁹ Second, official air quality data may be subject to manipulation by local governments (Chen et al., 2012; Ghanem and Zhang, 2014; Greenstone et al., 2020). Satellite-based measures seem more reliable: Their correlation with monitor-based PM_{2.5} data collected by US Consulates in China is approximately 0.8. City-level annual PM_{2.5} concentrations are measured using the Global Annual PM_{2.5} Grids derived from satellite data by Van Donkelaar et al. (2016).¹⁰ This yields a comprehensive air quality measure for a wide range of cities in China, covering all the prefecture, sub-provincial, and provincial cities.

⁸This is about 135 observations per city. A multistage cluster, stratified and PPS sampling methodology is adopted to ensure the national representativeness of the survey. As a result, the distribution of sample size across cities in CLDS is consistent with the geographic distribution of population in China.

⁹Fine particles (*diameter* < 2.5 μ m) are most hazardous and PM_{2.5} is considered to be the best indicator for health risks from air pollution. For more background information see WHO report: <http://www.who.int/mediacentre/news/releases/2014/air-quality/en/>. We use the Air Quality Index (AQI) released by Ministry of Environmental Protection (MEP) for robustness checks.

¹⁰They estimate ground-level PM_{2.5} by combining Aerosol Optical Depth (AOD) retrievals from the NASA MODIS, MISR, and SeaWiFS, which are subsequently calibrated to daily global ground-based observations of PM_{2.5} using Geographically Weighted Regressions (GWR) at high grid resolution.

3.3 Inputs into Instrumental Variables

We obtain information on large-scale (capacity > 1 million KW) power plants, their coal consumption, and plant-level electricity generation 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 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.

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. Data on the number of college students and graduates at city level are from the China Regional Statistical Yearbook. Second, we leverage variation in trade shocks to identify migration responses to wages. Data on Chinese trade are from the UC Davis Center of International Studies. The quantity and value of exports and imports by Harmonized System (HS) of product classification are available at the city level. Data are available annually between 1997 and 2013, covering periods before and after China's 2001 accession to the WTO. We construct city-level measures of baseline dependence on products more likely to be affected by tariff changes and trade policy.

3.4 Wages, Controls and City-level Characteristics

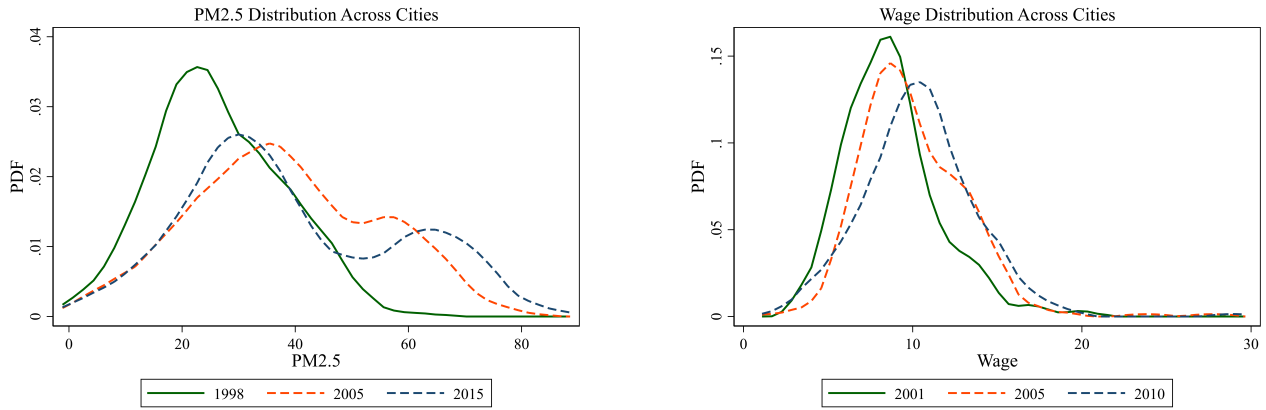
Since the 2015 Census does not record individual-level wages, we use the CLDS to calculate city-and-education specific average wage. We collect city characteristics, such as population and GDP, from the City Statistical Yearbooks. Weather data come from the Meteorological Data Service Center. We gather monthly data on temperature, humidity, sunshine duration, and other weather amenities. We calculate distances from each city to the three large seaports (Tianjin, Shanghai and Shenzhen) and employ these variables as controls. Appendix Table C1 reports summary statistics and a full description of the key variables used in the analysis.

4 Descriptive Patterns of Pollution and Migration

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

Figure 1 shows the increases in both city-level spatial dispersion of PM2.5 and real wages over time. The left panel documents that pollution not only increased between 1998 and 2015, but also became more variable across regions. The double-peak in the figure further indicates that the overall

Figure 1: The Distribution in Pollution and Wages Across Cities



(a) Increasing Spread in Pollution Across Cities

(b) Real Wage Distribution Across Cities

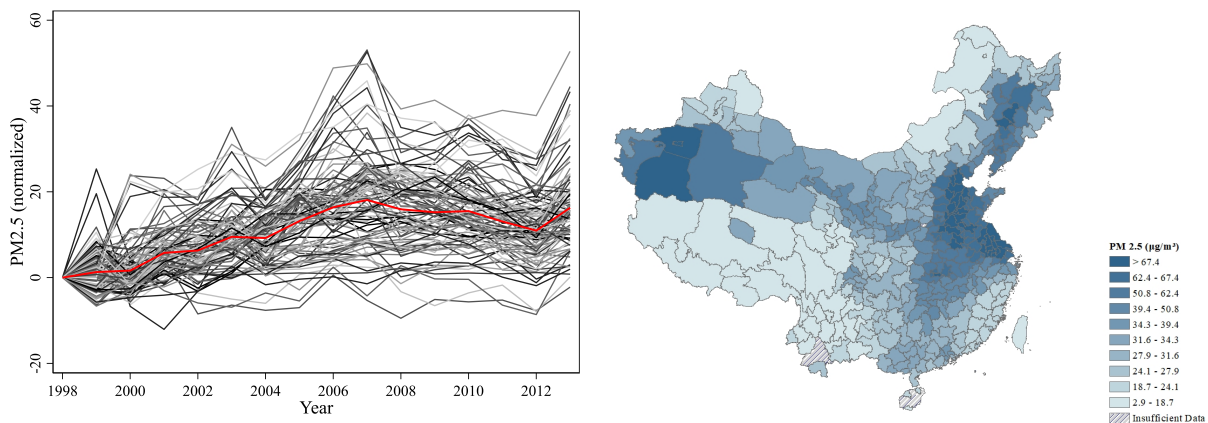
Notes: Distributions across cities for different years. Wage distribution drawn from the City Statistical Yearbooks. Real wages are nominal wages deflated by local housing price. PM2.5 data from the Global Annual PM2.5 grids.

increase in PM2.5 was driven by the emergence of some high polluting cities. The right panel shows that both wages and the cross-city variance in real wages rise over time. If this implies an increase in the spatial dispersion of marginal products of labor, then it raises the possibility that moving workers from low marginal product to high marginal product cities increases aggregate output.

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

(a) Change in mean PM2.5 over time

(b) PM2.5 Concentration in China



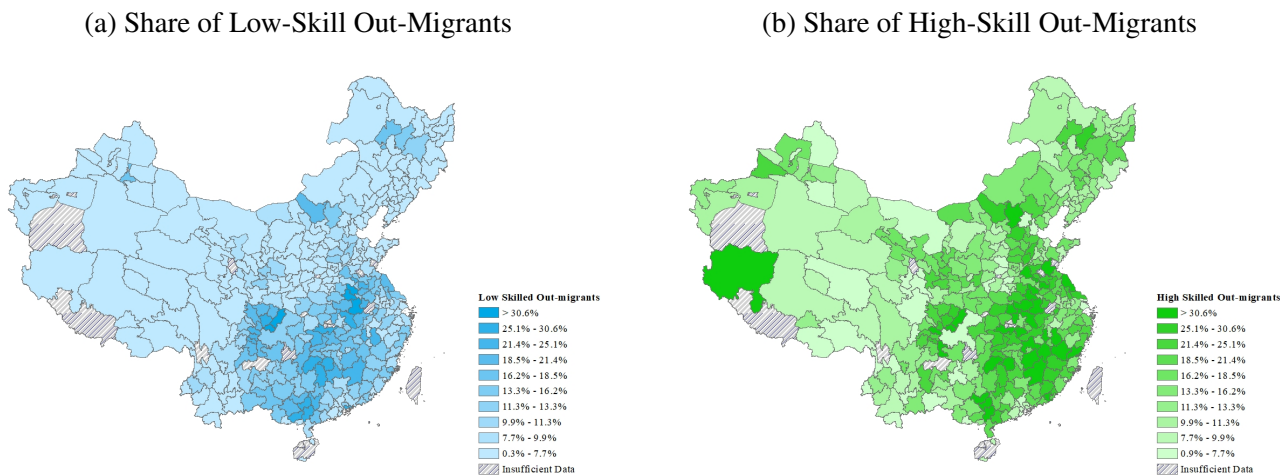
Notes: Spatial and temporal distribution of PM2.5 using the Global Annual PM2.5 Grids. The map shows the geographic spread in 2015. Figure 2a shows the increase in PM2.5 over time for the 100 largest cities in China, relative to the 1998 PM2.5 value (the difference with respect to 1998). The red line is the unweighted average.

Figure 2a illustrates the time trend of annual PM2.5 concentrations in Chinese cities since 1998. The mean concentration exceeds WHO air quality guidelines every year.¹¹ The figure also shows that the increase in the mean coincided with increases in cross-city dispersion of pollution (Figure 1a). The increase in the overall mean was driven by dramatic increases in PM2.5 in a subset of cities.

¹¹ See <http://www.who.int/mediacentre/factsheets/fs313/en/> for information on WHO guidelines.

Figure 2b shows the spatial variation in annual average satellite PM2.5 concentration for 2015. Air quality is unevenly distributed. The coastal north-east and east experience severe air pollution. Manufacturing industries concentrated in the east. The north-east further suffers from coal-burning for heating needs, exacerbating pollution relative to high growth areas of the south.¹²

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



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

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

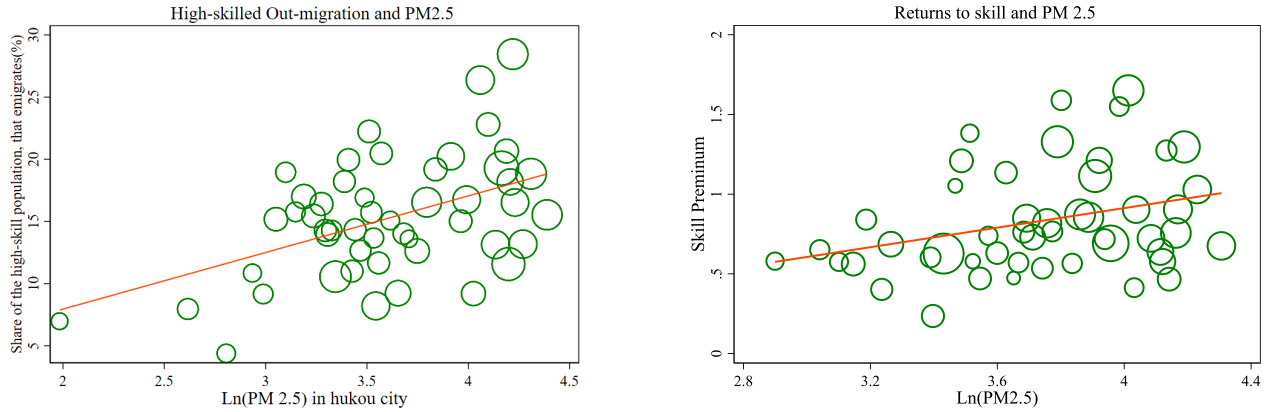
Figure 4a explores whether that observed spatial correlation creates a city-level association between pollution and the share of high-skilled that choose to emigrate. There is a clear positive association: the high-skilled are more likely to leave polluted areas. We explore this intriguing correlation more rigorously in subsequent sections, to isolate a causal relationship.

Finally, Figure 4b examines the association between pollution and the wage returns to skills that emerges in each city.¹³ Returns to skill are higher in polluted cities. Economic theory provides a simple explanation for the two related correlations depicted in Figure 4: Higher out-migration of college workers in response to pollution makes the high-skilled relatively scarce in those cities, and in equilibrium, creates a compensating differential for poor air quality for skilled workers. This relationship will endogenously emerge in the general equilibrium model of pollution, migration and wages that we develop. Figure 4b also highlights a key insight about the benefits of pollution control

¹²Dust-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.

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

Figure 4: The Effects of High PM2.5 at Origin Cities



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

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

Notes: The share of high-skilled out-migrants denotes the share of some college (or above)-educated out-migrants from the city-level 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.

policy that will emerge in our model: Reducing PM2.5 in high-polluted cities induce high-skilled workers to move to cities where their skills are relatively scarce (and so their marginal product may be relatively higher), and this re-sorting could raise aggregate productivity.

5 Identifying the Causal Effect of Pollution on Migration

Our data allow us to examine pairwise flows between cities, as we do when we estimate our labor supply curves in Section 8.2. First, we show the simplest specification, which studies the effects of PM2.5 concentration in origin city o on the amount of out-migration by skill group:

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

where M_{io} is an indicator for whether or not individual i left origin city o , and \mathbf{X} are controls. That is, we begin our reduced-form analysis with the dependent variable defined as the actual migration decision. Later, we show results using the net stock of workers by skill group, recognizing that 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.¹⁴

Before 2012, information about local PM2.5 concentration was not available in most Chinese cities. The Chinese government started releasing PM2.5 data in 2012, and that unexpected data disclosure affected the avoidance behavior of Chinese citizens (Jia et al., 2019). We therefore use the most recent census data in China – the 2015 One-Percent Population Census, conducted after the 2012 data disclosure – in our preferred specification. Indeed, the disclosure of pollution information

¹⁴As other work has noted, in such settings 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. As such, population quantities in all cities are affected, violating the Stable Unit Treatment Value Assumption (SUTVA). A GE model, like the one we posit, helps properly interpret the magnitudes in such a setting (Bryan and Morten, 2019).

itself will have a meaningful effect on migration, and in Appendix B we show that this disclosure affects both internal migration and air purifier purchases in China.

For many specifications, our primary variation is cross-sectional (across cities) so as to not be confounded by non-stationary trends in migration patterns and pollution level. In robustness checks, we also estimate this relationship using individual-level longitudinal panels with individual fixed effects.

Pollution is likely associated with the underlying structure of the economy, as polluted areas may have bigger manufacturing-bases, which independently affects the skill composition of the workforce. Furthermore, certain cities may be targeted by governments to improve air quality (see, for example, Beijing’s Clean Air Action Plan 2013-17), or to attract more skilled workers in less populated regions.¹⁵ OLS estimates of equation 1 may be biased as a result. To get around these issues, 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. To build confidence in our causal effect of pollution on emigration, our strategy is to examine the same relationship using multiple data sources, different types of variation, and conduct large sets of robustness checks (see Appendix A), to examine whether we see consistent patterns.

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

Our first source of plausibly exogenous variation in pollution is based on an insight from Freeman, Liang, Song, and Timmins (2019). We measure the extent to which each city is down-wind of a coal-fired power plant located outside the city. 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. The instrument value is penalized if the plant is farther away, and if it is not located directly upwind of the city, using the formula of our first-stage equation:

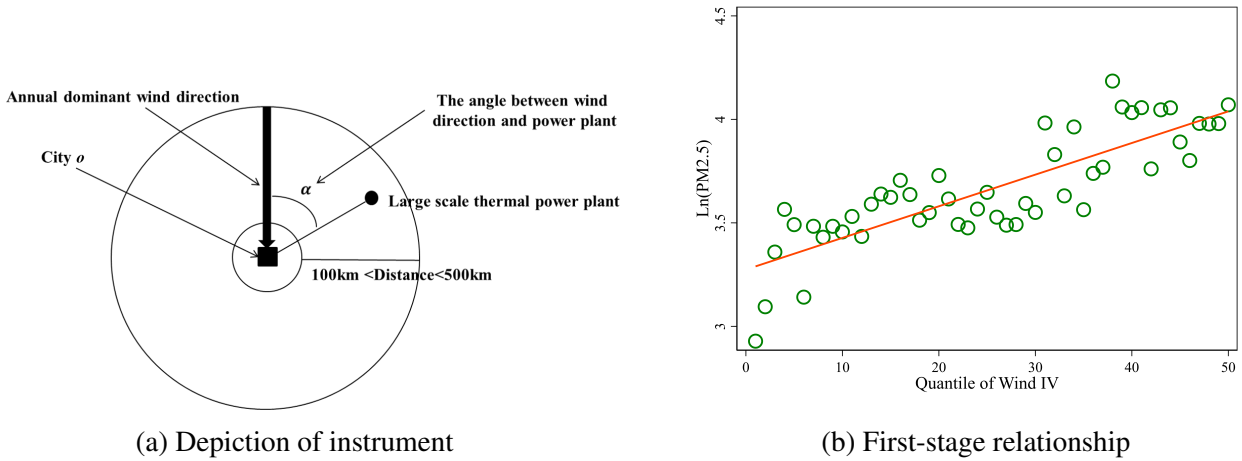
$$\text{Log}(PM2.5)_o = \gamma_0 + \gamma_1 \sum_p \left(\frac{1}{\alpha_{po} + 1} \right) \left(\frac{1}{\text{dist}_{po}} \right) C_p + \varepsilon_o, \quad (2)$$

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 . We restrict our analysis to all large-scale thermal power plants that are located more than 100km from a city, but within a 500km radius from the city center. Figure 5a explains the intuition behind the instrument. The underlying variation is driven by how wind patterns blow pollutants from distant coal plants to cities. Our first-stage relationship in Figure 5b shows that cities downwind from, and closer to, higher coal-consumption power plants are more likely to be affected by poor air quality.

We expect that our instrumental variable is orthogonal to local economic activity. These large-scale thermal power plants supply electricity to vast areas of China; many do not even supply electricity to their nearby cities, but rather to many remote provinces. The allocation of electricity supply from large-scale power plants is determined by the central government. Although many reforms have

¹⁵Our framework allows us to also examine whether certain cities are ‘too large’ (in which case, emigration would be welfare enhancing). Our results confirm Au and Henderson (2006) who show that Chinese cities are ‘undersized’.

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



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 fifty groups according to the quantile of the wind direction IV measure. The y-axis denotes the mean value of PM2.5 in each quantile and x-axis denotes the mean value of wind direction IV in each quantile.

taken place over the past 30 years, there are still strict regulations in the power sector and ownership of the sector largely remains with the state. The central government owns the grid, and controls the setup and operation of power plants if their generating capacity is large. Thus, local governments find it difficult to exert influence on the setup of large-scale power plants, their siting decisions, or allocation of electricity supply from them. The impact of distant power plants on local economic activity is extremely small, and we also chose to focus on plants at least 100km away to be cautious.

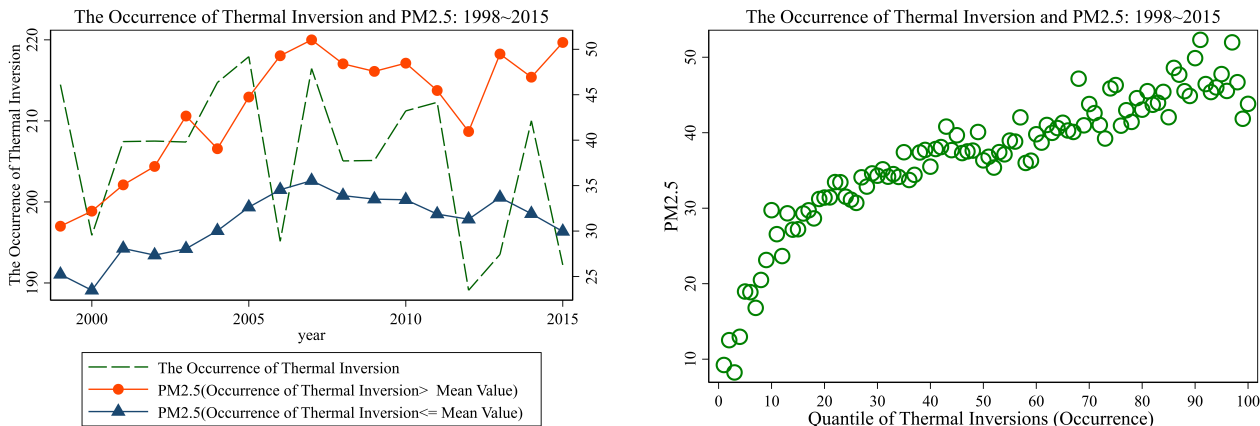
We examine possible concerns with this instrument in Appendix A.1.2. We consider whether the location of power plants depends on the simultaneous combination of wind direction, distance to the city, and the plant capacity. For instance, newly built plants may be placed away from important cities, and so we show robustness to excluding power plants in a 200km, and then a 400km radius from cities. Among other specification tests, we show robustness to excluding richer or capital cities, coal-producing regions, using other measures of air quality, and additional controls for electrification, demographics and industrialization. The public’s recent concerns around pollution are relatively recent, and we show robustness to using only old power plants, such as those built more than twenty years prior to our data. We also conduct numerous falsification tests showing that baseline city characteristics do not predict the future placement of plants, and tests with placebo wind directions indicate that downwind plants are less likely to affect air quality and migration.

5.2 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), the US (Hicks et al., 2015) and Sweden (Jans et al., 2014), among other places. Most recently,

Chen et al. (2017) show that the number of thermal inversions predicts the movement of people across China as well. We build upon their work by using newer migration data from the 2015 Census at the individual level, rather than quantifying migration indirectly from population changes.

Figure 6: Thermal Inversions and Air Quality



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

We create two measures of inversions in city o . First, we count the number of thermal inversions in each year. Next, we measure the annual mean strength of these inversions. We use these measures in both the cross-section and panel form.

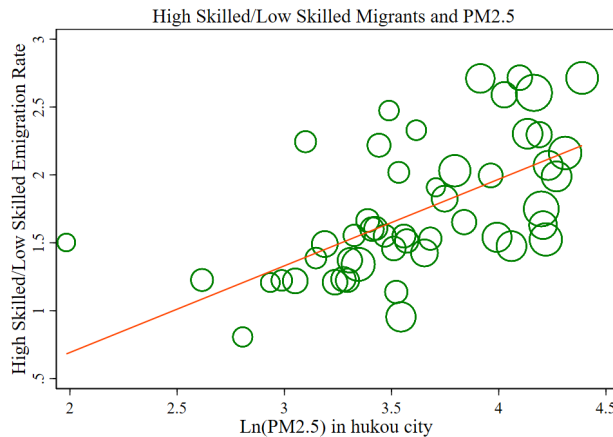
As polluting potential rose over time in China, areas with more thermal inversions trapped pollutants in the nearby atmosphere. Figure 6 shows that although there is no trend over time in the frequency of thermal inversions, as cities pollute more over time, areas with more inversions saw a sharper rise in poor air quality over the decade. The right panel shows a strong correlation between inversions and PM2.5 indicating a strong first stage. In our specifications we control for time-varying natural amenities like sunshine and weather, show variants of our measures of inversions, and show how past pollution does not predict future inversions.

We use the two instruments described above independently as they capture very different types of variation in air quality (due to wind direction versus a meteorological phenomenon). The consistency in the estimated direction of the relationship between pollution and migration is comforting. In specification checks in Appendix A, we compare the different instruments and their combinations, lagged panel structures, cumulative pollution over time, relevant controls, and excluding some cities. Finally, in Appendix A.2 we introduce yet another source of variation in pollution – a Regression Discontinuity (RD) across the Huai-river (Chen et al., 2013).

6 Empirical Results

Our theoretical model will help quantify the productivity consequences of pollution-induced migration. First, however, we need to establish that pollution indeed has a causal effect on (a) high and low-skilled differential migration rates, and (b) on the returns to skill in a city, because those are key inputs on which the model rests. We discuss reduced-form estimates of these relationships here, but relegate the more extensive discussion to Appendix A, which includes a long set of specification tests, falsification tests, sample restrictions, different panel-data structures, and different types of controls.

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



Note: High/low skilled emigration rates is the ratio of out-migration rates for high-skilled (some college or above) to low-skilled emigrants (high school or below). We plot these ratios in 2015 against PM2.5 levels.

6.1 The Causal Relationship Between Pollution and Migration

Figure 7 captures the relationship between differential mobility and pollution in the raw data. The ratio of high-skilled to low-skilled emigration rates increases with higher PM2.5 levels in the city of origin, suggesting that the location choices of high-skilled workers are more sensitive to pollution than that of low-skilled workers.

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 in China.

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 ($p\text{-val} < 0.001$). Columns 4-6 employ our first instrument based on wind direction and distant coal-fired power plants (Freeman, Liang, Song, and Timmins, 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

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*** (0.00974)	0.0399*** (0.00993)	0.0501*** (0.0112)	0.0772** (0.0389)	0.0609 (0.0423)	0.140*** (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*** (0.0202)	0.0871*** (0.0211)	0.112*** (0.0233)	0.0779*** (0.0228)	0.0740*** (0.0238)	0.107*** (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 equality of coefficients between the low and high educated groups, using the Fisher’s permutation test. 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.

(1.4 percentage points) than those without (0.61 percentage points). The high and low-skill emigration responses are statistically different from each other with >99% confidence.

In Panel B, we study how variation in PM2.5 from our second instrument of thermal inversions affects out-migration propensities. We follow the literature, and control for weather amenities. In the first three columns, we employ the annual occurrence of thermal inversions as the instrumental variable. 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 high education attainment, but only 0.87 for those without. In the last three columns, we leverage the variation coming from the annual strength of thermal inversions. We obtain similar results.

Appendix A.1.1 provides further details on our instrumental variable strategies. Table A1 shows (strong) first-stages. Table A2 estimates a stacked regression, interacting the pollution levels with skill-level indicators, and showing again that the effect on the high-skilled is statistically different

from that on the low-skilled. Table A3 shows similar results omitting weather controls, and Table A4 combines multiple instruments in the same specification, where we obtain even greater precision.

Table A5 shows that workers also pay attention to the pollution levels at destinations, in deciding which city to migrate to. In other words, workers leave polluted areas, and also seek out less polluted cities. Again, similar to the emigration response, the immigration location choices are more sensitive to pollution among higher educated workers. As destination choices are not independent across cities, we derive a model that disciplines these results. These facts play an important role in our model: Fewer skilled workers in a polluted city will tend to raise the skill premium. As a result, cleaning up polluted cities will move skilled workers from a city that has a lower marginal product of skilled labor to a city with a higher marginal product, increasing aggregate output via relocation.

In Section 8.2, we estimate a more comprehensive specification, documenting the pairwise flows between cities, and once again instrumenting for pollution levels with these very sources of variation. We estimate this relationship after our model, as it better allows us to describe what we are estimating when we allow for, say, changes in pollution levels in response to migration. Indeed, migration itself may affect both the levels and amounts of pollution, and so a more comprehensive model-based framework is required to understand these relationships.

6.2 Alternative Specifications, Robustness and Heterogeneity

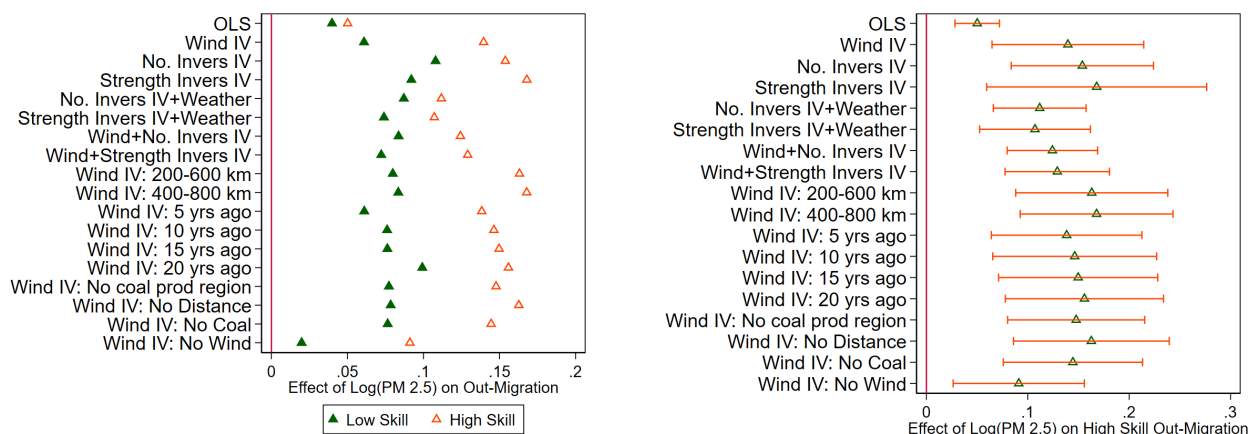
In Appendix A.1.2 we examine various threats to identification. We summarize the results of these exercises in Figure 8, where a simple pattern emerges: the high-skilled respond to poor air quality by emigrating, but the effect on the low-skilled is smaller.

First, we seriously evaluate the claim that power plants may be systematically built near poorer, less influential cities, and so the instrument may be correlated with unobservable characteristics of nearby cities. In Table A6, we exclude nearby plants, placed either in a 200km or a 400km radius of a given city, and find that, if anything, our results are more precisely estimated. We may also be concerned that newly built plants are subject to more regulation as the Chinese government only recently paid attention to environmental issues. So, in Table A7, we restrict our sample to only old plants, and find similar effects. We drop coal producing regions to address the concern that coal-fired plants may locate near coal production (Table A8), and show that baseline population, GDP and electricity consumption do not predict future upwind plants, or future iterations of our IV (Table A11). The results suggest that it is not that policymakers avoid richer, influential cities when building plants, and that plants are not built in areas that have a higher need for electricity at baseline, perhaps as most electricity is directly supplied to the larger grid.

In Table A12, we extend this by creating various ‘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 A13 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.2 we exploit an entirely different source of variation in pollution, driven by China’s

Figure 8: Different Sources of Variation



Notes: Summary of results using different sources of variation. We compile coefficients from different specifications. On the left we show both the coefficients on high and low skilled workers. On the right, we concentrate on high-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”). “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.

Huai river heating policy. As [Chen et al. \(2013\)](#) show, the heating policy generated an artificial discontinuity in air quality on two sides of the Huai river. North of the Huai river, the government encouraged centralized heating systems for homes, which primarily relied on coal-fired boilers. Even in the 2000s, there is a sharp discontinuity in the use of boilers for heating, leading to a discontinuity in air quality across the Huai river. While there is no differential out-migration in areas immediately north versus south of the Huai river, skilled workers are discontinuously (and statistically significantly) less likely to migrate into the more polluted cities just north of the Huai river, relative to cleaner cities just south of the river.

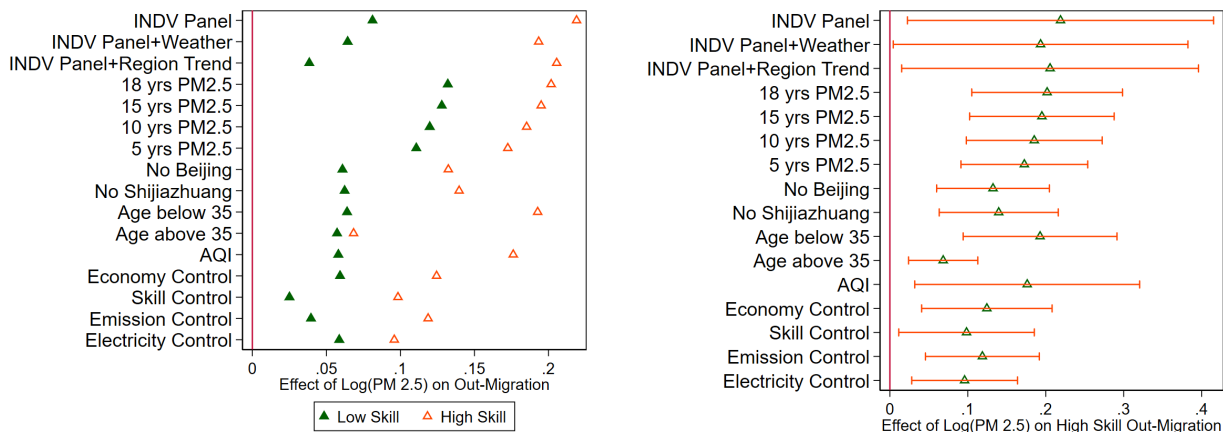
In [Appendix A.3](#), we turn our attention to studying different model specifications, sub-samples, and checking the robustness of our estimates to different controls. [Figure 9](#) summarizes the results of this large set of robustness exercises.

First, in [Appendix A.3.1 Table A14](#), we use individual-level longitudinal panel data and a different definition of migration to replicate our results. The longitudinal panel allows us to track individuals’ spatial sorting over time, and control for city and individual-level characteristics. Importantly, we define migration to be an indicator for whether or not an individual changed their city between years, regardless of whether they change their *hukou* location. This panel data with individual fixed effects approach produces the same pattern of results as the OLS and instrumental variables approaches: High-skilled workers are more likely to leave polluted cities relative to their low-skilled counterparts.

We then study the effects of cumulative pollution exposure in [Appendix A.3.2](#). The results shown in [Table A15](#) and [A16](#) indicate that pollution exposure spread over a longer time period has a larger impact than shorter time frames. In [Appendix A.3.3](#), we disaggregate education levels into more fine-

grained categories and see a sharp education gradient in out-migration responses. Those with more education are more responsive (Table A17). In Table A18 we exclude large, influential cities, cities that pollute a lot, and major province capitals, to account for any differences in political influence or outliers in the access to skilled jobs. We find that youth are more responsive to pollution when making location choices (Table A19). Table A20 studies heterogeneity across rural-urban locations.

Figure 9: Different Models, Samples and Controls



Notes: Summary of results using different models, samples and controls. We compile coefficients from different specifications. On the left we show both the coefficients on high and low skill workers. On the right, we concentrate on high-skilled workers, and include 95% confidence intervals. Individual longitudinal panel accounts for individual fixed effects (“INDV panel”), and for local weather conditions (“INDV panel+weather”) or region-specific trends (“INDV panel+region trend”)

We next replace PM2.5 with an alternative measure of local air quality. Table A21 uses the Air Quality Index (AQI) as the endogenous variable of interest, capturing the combined impact of many pollutants. The AQI is officially reported and widely disseminated, and individuals may be more likely to respond to it. In Appendix A.3.4 Table A22, other than the usual weather controls, we also control for annual rainfall, average slope (elevation) of the city, fraction of the region area under water coverage, and whether on the coast. Table A24 highlights robustness to a long list of socioeconomic controls, including the skill-distribution, baseline economic indicators, and industrial pollutants. These controls do not qualitatively affect the main patterns we observe.

Finally, in Appendix B, we show that the disclosure of official pollution data affected migration decisions. We look at how the elasticity between pollution and migration changes as data for local air quality was released. The releases were phased in, with the largest ramp up in 2012. Table B1 shows that the releases increased the migration elasticity for high-skill workers.

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.

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)	0.248** (0.125)	0.453*** (0.135)	0.789** (0.309)	1.570*** (0.539)
Observations	130	130	130	130
R-squared	0.028	0.090		
City controls	No	Yes	No	Yes
F-test of IVs			44.11	23.31

Notes: City-level regressions of 130 cities weighted by the population in 2000, for the sample of CLDS cities with non-missing skill-specific wage data. Robust standard errors are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai, 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 \(2002\)](#).

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

Table 2 documents a simple empirical fact: As in the raw-data plot of Figure 4b, returns to skill are higher in polluted cities. We estimate the city-specific Mincerian returns to skill using the CLDS data (as in [Dahl \(2002\)](#)), and explore the relationship between pollution and skill returns. In Section 8.1, Table 3 we look at the simple raw skill premiums as well. The first two columns of Table 2 show the OLS results. The impacts of pollution on skill returns are positive and statistically different from zero. The effect of pollution on the skill premium increases in magnitude when we employ the wind direction and coal-fired plants IV to address endogeneity concerns. Our model will formalize this link between pollution, differential out-migration of skilled workers, and the relative marginal products of skilled and unskilled work.

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

For the policy analysis we aim to conduct, it is important to understand precisely why skilled workers are more willing and able to leave polluted cities than the unskilled. Other than the typical pecuniary costs of migration, Chinese households must deal with institutional restrictions to mobility. Administrative records indicate that China’s *hukou* policy explicitly makes it easier for skilled workers to be mobile. People with higher education are more easily able to move to certain cities and retain

access to jobs, public services and schools. In Table C2 we show a few examples of such restrictions, highlighting how educational attainment can help a person gain enough points to be eligible to receive *hukou* at certain attractive destinations.

When migration costs (both physical and institutional) make it more difficult for certain subsets of workers to move, they may create artificial mismatches in the locations of skilled and unskilled workers who would otherwise complement each other in production. That can exacerbate productivity losses from worker re-allocation. The effect of poor air quality may also differ by skill level, as richer households may be able to afford air filters.

Beyond mobility costs, different preferences may also be important drivers of the differential mobility responses we observe. These include differences in preferences or attachment to certain cities based on the quality of amenities or schools, that may differ by skill level. Importantly, pollution concerns may loom larger for skilled workers, as the unskilled are focused more on making ends meet with their lower wages. Table C3 documents, using data from the China General Social Survey (CGSS), that educated workers are more likely to claim environmental issues in China are “terrible.” In Table C4, we use the survey to explore not just concerns for environmental issues, but also actions taken on such issues. Once again, across the board, individuals with more education are statistically significantly more likely to discuss environmental issues, make donations for environmental protection, and make appeals or raise concerns on environmental problems. The differences in migration patterns of the high-skilled in response to changes in pollution therefore partly reflect differences in preferences for environmental quality, and partly differences in mobility restrictions.

7 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, it endogenizes 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 level. 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 rise. If there is a degree of complementarity in production, the unskilled wage should fall, leading to differences in the skill-wage premium, consistent with the empirical observation that returns to skill are higher in more polluted regions. An additional source of geography-specific returns to skill is driven by the fact that some cities have more skill-biased capital.

Furthermore, the changing structure of skills in a city affect production and pollution levels. A population influx, or presence of more skilled workers may induce either more or less pollution-intensive industries to expand, and as such change the quality of air in the city. This feedback effect of migration patterns on pollution emissions affects subsequent migration, which in turn affects production, and so on. Additionally, agglomeration forces may increase aggregate productivity if skilled workers converge to high amenity cities, but house prices may also respond to such movements cre-

ating congestion in such cities. Importantly, we allow pollution to affect the health (and lower the productivity) of all workers.

Our framework accounts for these feedback effects, and generates estimable equations that we identify using instrumental variables. We use the model to perform quantitative counterfactuals on the equilibrium effects of pollution control policies.

7.1 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. Effective labor L_d depends on labor L_{sd} at each level $s = \{h, u\}$, the high-skilled h and the unskilled u .

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

$0 < \rho < 1$ 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 a degree of 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 (Appendix D.2). 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 k_{hd} , such that $\theta'_{hd}(k_{hd}) \geq 0$.¹⁶ 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 k_{sd} , 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:¹⁷

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

where $\log \tilde{\rho} \equiv \left(\frac{1-\rho}{\rho} \right) \log \left(\frac{1-\rho}{R^*} \right)$ 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.¹⁸ There are a few components that drive the differences in average earnings when comparing two different skill-groups s in two different labor markets d .

Let us consider the determinants of the high-skilled wage w_{hd} . First, A_d is the amount of TFP at the

¹⁶Skill-biased capital for high-skilled workers captures the presence of industries that may be more likely to hire college-educated workers, such as finance, technology, skill-intensive manufacturing, and professional services. For completeness, in Appendix D.2 we explicitly model skill-biased capital within the nested CES framework and show how incorporating it does not affect the qualitative predictions.

¹⁷This is at the optimal value of K_d^* , so that $Y_d = A_d^{\frac{1}{\rho}} \left(\frac{1-\rho}{R^*} \right)^{\frac{1-\rho}{\rho}} L_d$.

¹⁸We keep our focus on migratory frictions, and abstract away from city-specific trade costs. If trade was not frictionless, 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.

city level, which raises average earnings of all in the city. 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 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 by skill. First, we allow economic production and the location of workers affects pollution as well. In equation 5, each city has some ‘exogenous’ component of pollution \bar{Z}_d , that (say) depends on the occurrence of thermal inversions. Yet, the equilibrium pollution levels Z_d depend on other endogenous factors as well. As the workforce composition changes, some cities produce more output than others, which in turn raises pollution. The increase in pollution depends on both aggregate population (the size of the economy, industrial production and congestion) (Rendon et al., 2014), and the skill mix. The skill mix affects pollution by changing the type of production (industry vs services), amenities, energy consumption, or even local pollution policies:

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

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. We expect the number of skilled workers to raise TFP levels in the city via non-excludable innovation (Arrow, 1962), such that $\phi_2 \geq 0$:

$$A_d^{\frac{1}{\rho}} = \bar{A}_d Z_d^{\phi_1} L_{hd}^{\phi_2} \quad (6)$$

7.2 Migration and Labor Supply

We assume that workers have preferences over locations, either because of tastes (some prefer to be closer to home, others prefer big cities), or because it is more ‘costly’ for some to migrate and leave home. The indirect utility of worker j , with skill group s , in destination d , from origin city o is:

$$V_{jsod} = \varepsilon_{jsd} w_{sd} Z_d^{-\gamma_s} h p_d^{-v_s} a_{sd} \exp^{-M_{sod}}, \quad (7)$$

where ε_{jsd} is a random variable measuring preferences for a specific city d by individual j . A larger ε_{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. As such, migration costs vary by education level, and $M_{s0o} = 0$ for natives. hp_d are housing prices, and v_s are the share of expenditures on housing by skill-level. a_{sd} represents other non-pollution related skill-specific amenities. The compensating differential elasticity will be captured by γ_s , and varies by skill.¹⁹ Here, marginal workers are indifferent across cities, and so are likely to be induced into migration by pollution. Inframarginal workers have higher utility in the city they live in currently, than in other cities.

We assume that ε_{jsd} are independently distributed and drawn from a multivariate extreme value distribution (Eaton and Kortum, 2002). The joint distribution of ε_{jsd} is given by:

$$F(\varepsilon_{s1}, \dots, \varepsilon_{sD}) = \exp\left(-\sum_d \varepsilon_{sd}^{-\eta_s}\right), \quad (8)$$

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

Workers move to places where their utility is higher, implying that given costs of moving, there are no further arbitrage opportunities available. Local ties and migration costs (including *hukou* and distance-based costs) are captured by ε_{jsd} and M_{sod} respectively. A person chooses city d over d' if $V_{jsod} > V_{jsod'}$. In Appendix D.1 we derive an expression for labor supply and average city-utility, where the share of workers with skill s from city o that move to d is:

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

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

$$L_{sd} = \sum_o P_{os}\pi_{sod} \quad (10)$$

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

$$\log \pi_{sod} = -\eta_s \log \bar{V}_{so} + \eta_s (\log w_{sd} - v_s \log hp_d) + \eta_s \log a_{sd} - \eta_s \gamma_s \log Z_d - \eta_s M_{sod}, \quad (11)$$

where \bar{V}_{so} captures the average utility of being from city o .²⁰ 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

¹⁹If $\gamma_h > \gamma_u$, good air quality is a normal good.

²⁰As we show in Appendix D.1, we can derive: $\bar{V}_{so} = \left(\sum_{d'}\left(w_{sd'}Z_{d'}^{-\gamma_s}hp_{d'}^{-v_s}a_{sd'}\exp^{-M_{sod'}}\right)^{\eta_s}\right)^{\frac{1}{\eta_s}}$.

location falls. Therefore, as we show in Appendix D.1, \bar{V}_{so} depends on city-specific mobility costs.

From equation 11 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. As such, 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.

7.3 Equilibrium and Elasticities

Equations 3-11 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 to reside in. To be specific, the model is characterized by a set of exogenous factors: city level productivities \bar{A}_d , total 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 together determine the quantities $A_d, Y_d, L_{hd}, L_{ud}, Z_d$ and prices w_{hd}, w_{ud}, hp_d .

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 $\{h, u\}$, as in equations 4, 9 and 10. Total population of the skilled and unskilled in the country is the sum of the city level populations (or, $\bar{L}_h = \sum_d L_{hd}$ and $\bar{L}_u = \sum_d L_{ud}$). The sum of shares of migrant and non-migrants adds up 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. Aggregate output is simply the sum of output in each city $Y = \sum_d^D Y_d$. In Appendix D.3 we describe conditions that determine the existence and uniqueness of the equilibrium.

8 Estimation of Model Parameters

Section 6 already confirmed two important model results. As (exogenous) air pollution increases, either skilled labor emigrates out or fewer skilled immigrate in, raising the returns to skill such that the difference in wages is the compensating differential. 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 (which determines the compensating differential). Instead, in general equilibrium it is the result of corresponding migration changes as wages change, given η_s (the labor supply elasticity based on local ties and preferences). The wage changes in turn depend on σ_E , the elasticity of substitution across skill-levels in production (or the 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 many model parameters. These crucial elasticities determine the productivity consequences of pollution, which we need to estimate.

We estimate the aggregate 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. In our preferred estimates we control for city characteristics as before, and show robustness to alternative sources of variation. We find estimates of these parameters to be similar to what has already been estimated in the literature. While we could calibrate our model from parameters estimated by other researchers, we feel more confident in using parameters estimated within our own model and data. To generate causal estimates for these parameters, we exploit additional sources of variation stemming from other policy changes.

8.1 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-level characteristics) are differenced out, as in equation 12:

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

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:

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

We perform this exercise in Table 3, using each of our instruments for pollution in each column. We 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).

In Table 1 of Section 6, we had used individual-level data to show how the emigration response to pollution was larger for the high-skilled than the low-skilled, while Table A5 showed that in-migration was similarly skill-biased. These differential migration rates affect the relative quantities 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 the relationship in equation 12. We take the ratio of the IV relationship for quantities of workers, and the wages of workers. For instance, in the first two columns, we find that $\sigma_E = 1.238/1.000 = 1.238$. This suggests that the elasticity of substitution across skill levels is 1.238, an estimate close to the estimates found 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.

Table 3: Estimating Labor Demand Elasticities

IV:	Coal+Wind IV		No. of inversions		Inversion strength	
	Ln $\frac{w_h}{w_u}$ (1)	Ln $\frac{L_h}{L_u}$ (2)	Ln $\frac{w_h}{w_u}$ (3)	Ln $\frac{L_h}{L_u}$ (4)	Ln $\frac{w_h}{w_u}$ (5)	Ln $\frac{L_h}{L_u}$ (6)
Log(PM2.5)	1.000** (0.438)	-1.238** (0.617)	0.703*** (0.244)	-1.458*** (0.366)	0.850*** (0.316)	-1.681*** (0.554)
Observations	130	130	130	130	130	130
City Controls	Y	Y	Y	Y	Y	Y
Weather	N	N	Y	Y	Y	Y
F-stat	20.54	20.54	68.84	68.84	45.41	45.41
σ_E	1.24		2.07		1.98	

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 below. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Robust standard errors reported in parentheses. All regressions weighted by the population in 2000. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

The labor supply curve in equation 11 captures bilateral migration flows between pairs of cities as a function of real wages and pollution at possible destinations, and migration costs between origin-destination pairs. First, we parameterize migration costs M_{sod} as follows:

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

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 .²¹ 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 11 generates an estimable equation for labor supply:

$$\begin{aligned} \log \pi_{sod} = & -\eta_s \log \bar{V}_{so} + \eta_s (\log w_{sd} - v_s \log hp_d) - \eta_s \gamma_s \log Z_d + \mathbf{X} \beta_x \\ & - \eta_s \lambda_{1s} \log Dist_{od} - \eta_s \lambda_{2s} (\mathbb{1}_{Migrant_{od}} \times hukou_{sd}) + \epsilon_{3sod}, \end{aligned} \quad (15)$$

where the residual $\epsilon_{3sod} = (\eta_s \log a_{sd} + \epsilon_{2sod})$ includes differences in destination city amenities and other idiosyncratic features determining bilateral flows. Below, we describe how we derive amenities from the residual by inverting the model. In our estimation, we include origin-city-by-skill fixed effects to control for $\eta_s \log \bar{V}_{so}$. 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

²¹Specifically, the index measures the difficulty of obtaining a local *hukou* based on one's job, family reunion motives, local investments, and contribution to the city's workforce. A higher index indicates a more restrictive policy. The index is highest for Beijing, followed by the other major cities.

costs, while controlling for the index accounts for any differences in city level characteristics (correlated with the *hukou* index), that affect migrants and non-migrants in a similar fashion. Like before, we instrument for pollution using thermal inversions.

One of the key parameters of interest are labor supply elasticities η_s for skilled and unskilled workers. To estimate these elasticities, 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. \(2013\)](#) have studied the effects of these shifts on US manufacturing employment, but here we focus on their effects on skilled and unskilled wages in China.

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:

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

Comprehensive details about this trade shock can be found in [Khanna et al. \(2020\)](#), where they show that the NTR gap instrument better predicts changes to real unskilled wages, rather than skilled wages, possibly because the industries that benefited most from such tariff changes were more likely to hire unskilled labor. We use this as an instrument for unskilled real wage $\log w_{ud} - v_u \log hp_d$. We ‘deflate’ all our wages by local house prices.²²

For skilled wages we derive variation from the World Import Demand (WID) for skilled industries.²³ Following [Autor et al. \(2013\)](#), we use world-import demand shocks by industry, and weight them by initial export shares to derive city-level exposures. As we aim to create an instrument for skilled-wage, in equation 17 we use the share of skill-intensive industries:²⁴

²²We use yearly average data on housing rents from the Xitai Real Estate Big Data depository. Further details and data sources at <http://www.cityre.cn/credata.html>. The data has been collected since 2005, and covered 337 cities in China, in collaboration with the China National Bureau of Statistics, and the China National Development and Reform Commission. We compare these data to the purchase price of residential properties from the China statistical yearbook, and find a correlation of 0.93.

²³We label industries as skill intensive if they are above the median in the ISIC industry data. The skill share is the share of skilled workers in the industry, based on the Annual Survey of Industrial Production (ASIP) available in 2004. We aggregate the firm data into 4-digit ISIC industries. For instance, in ISIC 1810, 5% of the labor force is “skilled”. We construct our measure using the Indonesian manufacturing census ([Amiti and Freund, 2010](#)), so as to ensure no confounding effects of using the same sample to construct our skill-intensity measure and regression estimation.

²⁴[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. \(2018\)](#) and [Adao et al. \(2019\)](#)), rather than the ‘shares’ being exogenous ([Goldsmith-](#)

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

Table 4: Estimating Labor Supply Elasticities

IV-2SLS Labor Supply	Low Skill Workers		High Skill Workers	
	$\text{Log } \pi_{uod}$		$\text{Log } \pi_{hod}$	
	(1)	(2)	(3)	(4)
$\text{Log}(PM2.5)_d$	-0.0427*** (0.0090)	-0.0488*** (0.0114)	-0.506*** (0.0958)	-0.513*** (0.0829)
$\text{Log}(\text{Real Wage})_d$	1.012*** (0.269)	1.126*** (0.318)	1.301*** (0.251)	1.024*** (0.170)
$\text{Log}(\text{Distance})_{od}$	-0.0754*** (0.0112)	-0.0783*** (0.0129)	-0.0308*** (0.0052)	-0.0404*** (0.0046)
$\text{Hukou Index}_{sd} \times \text{Migrant}_{od}$	-0.923* (0.487)	-0.852* (0.501)	-3.489*** (0.876)	-3.061*** (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 NTR IV is the weighted average of the NTR gap, where the weights are the baseline industry level export shares (Pierce and Schott, 2016). The NTR gap is measured as the gap in Normal Trade Relation (NTR) tariffs and the non-NTR tariffs. The WID IV is the weighted average of the world import demand, where the weights are the baseline skill-intensive share of industries Autor et al. (2013). The measure of *Hukou Index_{sd}* varies across cities and skill level (Zhang et al., 2018). We also control for this (non-interacted) measure of the hukou index. We model distance as inverse hyperbolic sine, to be able to include zero values. All regressions also control for temperature, humidity, sunshine duration, and wind speed, as before when using thermal inversions as an IV. High skill workers are those whose highest degree is some college or above, and low skill workers are those whose highest degree is high school or below. Standard errors clustered at the city level are reported in parentheses. The first stage relationships are described in Table C5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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).

Equation 15 shows that the coefficients on $\text{Log}(PM2.5)_d$ equal $\eta_s \gamma_s$. The γ_s parameters capture the marginal utility of clean air, and vary by skill level. 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 than the unskilled. Comparing estimated labor supply

Pinkham et al., 2020).

elasticities with respect to wages and pollution levels suggests that both types of workers are far more responsive to changes in 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 access to amenities (like housing purchases and children’s schooling) only obtainable via accessing local *hukou*.

As workers move in response to higher wages, this affects where production takes place, and as a result, where pollution is located (based on ψ_1 and ψ_2). As such, our model and estimation allow for trade shocks to also affect the amount of pollution via production and migration responses.

8.3 Measuring Amenities and Productivities $\{\theta_{sd}, \alpha_{sd}, \bar{A}_d\}$

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

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

We plot the city-level distribution of θ_{hd} in Figure C2. Beijing, Shanghai and other urban centers have high amounts of high-skill capital than the less urbanized, less developed areas.

We follow the literature (Ahlfeldt et al., 2015; Bryan and Morten, 2019) and derive non-pollution amenities as a residual from the labor supply curve, equation 15. Intuitively, bilateral migration flows between cities must be driven by wages, house prices, pollution, migration costs, and other residential amenities. As we measure and account for the all other components, the remaining portion of flows are due to amenities. We take the residual for each skill-specific regression estimated from equation 15, and derive amenities from destination fixed effects (i.e., $\varepsilon_{3sod} = \alpha_{sd} + \varepsilon_{2sod}$, where $\widehat{\eta_s \log a_{sd}} = \widehat{\alpha_{sd}}$).

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. After accounting for the optimal (unbiased) capital flows, output is simply: $Y_d = A_d^{\frac{1}{\rho}} \left(\frac{1-\rho}{R^*} \right)^{\frac{1-\rho}{\rho}} L_d$. Using our estimates for σ_E and θ_{sd} , we are able to create a measure of L_d . We then invert the model to derive $A_d^{\frac{1}{\rho}}$.²⁵

²⁵ $\log \bar{\rho} \equiv \left[\left(\frac{1-\rho}{\rho} \right) \log \left(\frac{1-\rho}{R^*} \right) \right]$ is common across all cities and workers. ρ is the labor share of income.

Table 5: Pollution, Population and TFP

	Agglomeration Forces		Congestion Forces	
	Log(TFP) (1)	Log(TFP) (2)	Log(PM2.5) (3)	Log(House Prices) (4)
Log(PM2.5) _d	-0.0816 (0.256)	-0.0595 (0.285)		
Log(L _{hd})	0.0970* (0.0528)	0.0964* (0.0519)		
Log(Population) _d			0.266* (0.150)	0.259** (0.121)
Log(L _{hd} /L _{ud})			-0.00781 (0.0786)	0.423*** (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). The first-stage relationships are in Table C5. Robustness for house-price regressions are in Table A23. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

8.4 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.

$$\frac{1}{\rho} \widehat{\log A_d} = \log \bar{A}_d + \phi_1 \log Z_d + \phi_2 \log L_{hd} + \varepsilon_{4d}, \quad (19)$$

where ϕ_1 is the elasticity of pollution with aggregate TFP.²⁶ 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 Arrow (1962) style innovation spillovers, it would be captured by the agglomeration elasticity ϕ_2 .²⁷

In Table 5 we estimate ϕ_1 using equation 19, and leveraging our instrumental variables strategy for PM2.5. While not precisely estimated, we conclude that $\phi_1 = -0.0816$. Chang et al. (2019) document an elasticity of -0.023 for call-center workers in China, and Kahn and Li (2019) estimate an elasticity of -0.18 for public sector workers in China. Our estimate lies in between.

We use a sudden expansion of access to universities in China to estimate the effects of skilled workers on TFP, as a way to capture the parameters associated with agglomeration or innovation spillovers. The Chinese government instituted a policy to expand college enrollment in 1999, pri-

²⁶TFP may also be positively associated with pollution. This may generally be true if more production leads to more pollution, but we are using an IV for pollution here.

²⁷Agglomeration in this model is represented by the production of non-excludable ideas. Innovators are not directly compensated for ideas in wages. Instead overall output increases, benefiting all in the local economy.

marily by lowering the bar for admission.²⁸ Che and Zhang (2018) use this policy as an instrument for firm productivity, and discuss identification concerns at length. They establish that the university expansion policy was not targeted at any specific industrial sector. We use this same instrument to predict variation in the number of skilled workers across cities. Figure C1 describes this event graphically. Importantly, it shows no differential pre-trends in college students and professors (who subsequently make up the skilled workforce) in cities that subsequently benefited from college expansion. Furthermore, Table C6 suggests that, as expected, 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.²⁹ We find $\phi_2 = 0.097$ when estimating equation 19, which suggests meaningful agglomeration in line with estimates in the literature (Gaubert, 2018; Moretti, 2004; Peri et al., 2015).

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

Simultaneously, we estimate how changes in total population affect air quality, captured by parameter 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 out-migration between 2005 and 2015 from all provinces in China, weighted by the share of migrants from each province that came to d in 2000:³⁰

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

Note that this is different from the instruments introduced by Card (2001) and described by Jaeger et al. (2018). As we have rich data on outflows of migrants from provinces, we can leverage ‘push factors’ from sending regions, 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 city d specifically, similar to the shift-share framework described by Borusyak et al. (2018). We combine this idea with the strength of the Card (2001) framework, in which sources

²⁸In January 1999 the Ministry of Education (MOE) announced an admission plan of 1.3 million for three and four-year college programs, a 20% increase over 1998. The following June it revised the admission plan to 1.56 million, an unprecedented increase of 44% over the previous year. College admissions grew annually by more than 40% in both 1999 and 2000, and by about 20% over the next five years. The gross college enrollment rate among 18–22 year-olds increased from 9.8% in 1998 to 24.2% in 2009. The year 2003 saw the first flow of four-year graduates into the job market as a result of the expansion. The number of students graduating from regular higher education institutions was 2.12 million, a 46.2% increase from the previous year.

²⁹The instrument is number of college graduates in city d in 2005 minus the number of college graduates in city d in 2001. Figure C1 describes the dynamics of the expansion policy, and pre-trends.

³⁰When 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.

Table 6: Summary of Estimated Model Parameters

Parameter	Value	Definition	Identifying Variation
σ_E	1.24	Skill-elasticity of substitution Relative labor demand elasticity	Pollution-driven geographic sorting Changes in skill ratio affect wages
η_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	0.423	House price response to skill ratio	University expansion
ψ_4	0.259	House price response to population	Out-migration from origin provinces
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 C2. 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.

of migration at baseline determine the migrant networks that attract more migrants into city d after 2005, whenever there are larger outflows from the set of origin provinces that d was connected to.³¹

Table 5 describes the effect of population and skill ratio on pollution. The first stage of the 2SLS appears to be strong. 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 ($\psi_3 = 0.423$), and with respect to the skill ratio ($\psi_4 = 0.259$), 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 A23.

8.5 Model Solution and Validation Exercises

Table 6 summarizes all parameter values and sources of variation. 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: $\{Y_d, A_d, L_{hd}, L_{ud}, Z_d\}$, 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. 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 our endogenous quantities, like GDP and wages.

³¹Indeed, before 2000 internal migration was highly regulated (Kinnan et al., 2018). After 2000, migration costs have fallen substantially (Tombe and Zhu, 2019), allowing us to estimate the impacts of subsequent growth.

We solve the model starting with a list of parameters and exogenous quantities, and a set of initial conditions for the endogenous variables. After estimating the parameters and exogenous quantities, we no longer use the data to solve the model. The primary market clearing condition is the labor market equilibrium, since for tractability, we do not model the floor space market, or consumption. We pick different starting points, beginning in the vicinity of the observed equilibrium, but vary it by as much as changing the starting value by 20% for each endogenous variable. The model converges to the same unique equilibrium.³² We solve the model iteratively, ensuring that we clear the labor market. That is, if there is an excess supply of labor in a city, wages fall, and along with the population constraints (the total number of skilled and unskilled in the country) our model converges to an equilibrium determining the distribution of skilled and unskilled workers across cities.

We test for model fit in Figure C3, and show that our model’s predictions match the endogenous quantities in the data. We use our estimated parameters to calculate predicted values for all cities. We test model-fit by first plotting predicted $\text{Log}(GDP)$ and skill-premia $\text{Log}(w_h/w_u)$. In the lower panels we plot actual and predicted $\text{Log}(L_h/L_u)$ and L_h . Notice, these are not necessarily out-of-sample tests as we use the data when estimating the model parameters, but not when solving for equilibrium. Across all measures, our model replicates the major spatial data patterns.

9 Counterfactuals: The Gains from Relocating Pollution

There are three mechanisms through which pollution can result in productivity losses. First, and the one this paper seeks to highlight, skilled workers leave cities where they would be more productive, to avoid pollution. Further, since migration costs are larger for the unskilled and they are less responsive to pollution, the unskilled do not leave with the skilled. As we estimate the skilled and unskilled to be fairly complementary in production, this creates a mismatch in the location choices of the two types of workers, further reducing aggregate productivity. Second, pollution reduces the agglomeration of skilled workers in the most productive cities, lowering aggregate output. 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. 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 that they must meet. In the second type of policy, we relocate the exogenous component of pollution only, \bar{Z}_d . This policy is similar to relocating coal-fired plants from

³²This 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). 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. 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.

upwind to downwind regions. Note that relocating plants would induce new residents to move into the city, change production patterns, and a revised pollution level will emerge in a new equilibrium that will not necessarily equal the modeled change in \bar{Z}_d .

Relocating plants away from high-capital cities to low-capital cities, will benefit certain cities but hurt others. As such, even though the relocation may improve aggregate productivity, the distributional effects across cities and skill groups are more complex.

9.1 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 the steady state PM2.5 concentration 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 GDP per worker 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. For instance, the newly relocated immigrants also benefit from better health.

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 powerplant up-wind 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.

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

	Change in GDP per Worker in Beijing (%)			
	Overall Change (1)	Health Channel (2)	Relocation (3)	Relocation+Agglom (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.

In the third and fourth rows of Table 7, we examine the effects of relaxing *hukou* restrictions in Beijing by 50%, allowing workers to move in, but holding pollution fixed. The effects of relaxing

mobility restrictions depend on whether we change *hukou* policy for skilled or unskilled workers. When we relax the skilled *hukou*, GDP per worker rises by 8.2%, more than half of which is driven by the simple relocation channel, and the rest by the agglomeration forces. Lowering unskilled *hukou* restrictions lowers GDP per capita, through a compositional change in the population – there are now more low-wage workers in Beijing.

The last two rows combine the changes in steady-state pollution and *hukou* restrictions. Relaxing skilled *hukou* restrictions in Beijing increases city GDP per capita by 22.96%. 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%.³³

Relocation due to pollution or *hukou* policy changes 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.

Table 8: Distributional Consequences of Reducing Pollution in One City

	Skilled Wage in Beijing			Unskilled Wage in Beijing		
	Overall (1)	Health (2)	Relocate+Agglom (3)	Overall (4)	Health (5)	Relocate+Agglom (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: 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. 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.

We move beyond overall GDP per capita to examine distributional consequences on the wages of each skill group in Table 8. When we reduce the steady state amount of pollution in row 1, skilled wages rise slightly. The improved productivity of skilled workers from 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

³³ Appendix Table E1 examines changes to GDP per worker, focusing solely on the main relocation effects and shutting down agglomeration or congestion effects (endogenous changes to pollution, house prices and TFP).

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. For instance, relaxing skilled *hukou* lowers skilled wages by 4.3%, but as skilled workers enter the city, unskilled workers become more productive and their wages rise by 12.5%. When we combine relaxing *hukou* regulations with changes to PM2.5, there are more positive effects on wages, especially for the unskilled.

9.2 Relocating Pollution Away from Skill-biased Capital

Table 9 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 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 much 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.

Table 9: The Productivity Effect of Relocating Pollution Across Cities

	Change in GDP per Worker in China (%)			
	Overall changes (1)	Health (2)	Relocation (3)	Relocate+Agglom (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.

When we relocate the exogenous part of pollution (say, shift an upwind power plant away from a productive city, to a less productive one), the increase in GDP is smaller. As people relocate to the more productive cities, pollution levels again rise through the feedback loop in the model (where pollution is itself a function of population), and dampens the benefits. The contribution of the relocation channel is slightly larger than the health channel.³⁴

³⁴The health channel playing a positive role suggests that sources of pollution are concentrated in more populous cities.

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 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.

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 10: Distributional Effects of Relocating Pollution Across Cities

	Skilled Wage			Unskilled Wage		
	Overall (1)	Health (2)	Relocate+Agglom (3)	Overall (4)	Health (5)	Relocate+Agglom (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: In this counterfactual exercise we relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relocate only the exogenous component of pollution. In addition to such relocations of pollution, we also relax the *hukou* restriction in the 24 top tier cities by 50% (row 3). In row 4 we relax overall migration costs to the 24 high tier cities by 50%. Rows 3 and 4 keep pollution fixed. Columns 1-3 show the effects on skilled workers, while columns 4-6 show the effects on unskilled workers.

Table 10 shows wage effects for both the skilled and unskilled. 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. As skilled workers leave cities where pollution increases, skilled wages rise in those cities and wages of unskilled (complementary) workers fall. 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 they 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. Wages for the unskilled improve slightly by 1%.

Relaxing *hukou* restrictions and lower migration costs (rows 3 and 4) raise wages for both skilled

Moving pollution away from large cities can improve population-weighted average health.

and unskilled, as workers match better with where their marginal products are higher. That is, workers locate to where there is more capital, and where there are more complementary workers. 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.

Table 11: Welfare Effects of Relocating Pollution Across Cities

	Skilled Welfare (1)	Unskilled Welfare (2)	Average Welfare (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: In the counterfactual exercise we relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relocate only the exogenous component of pollution. In addition to such relocations, we also relax *hukou* restrictions in the 24 top tier cities by 50% (row 3). In row 4 we relax overall migration costs to the 24 high tier cities by 50%. Rows 3 and 4 keep pollution fixed. Rows 5 and 6 relocate pollution while relaxing migration costs.

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

Table 11 examines changes to welfare by skill. Relocating pollution away from cities with skill-biased capital raises the welfare of skilled workers by 29.7%, as it raises their wages, but also 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%.

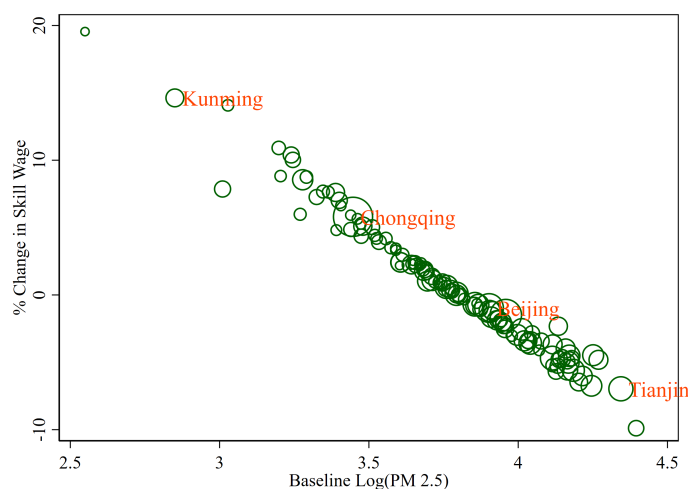
Lowering migration costs (row 4) raises the welfare of both the skilled and unskilled by 17.5% and 10.1% respectively. The combination of relocating pollution and lowering migration costs improves welfare in the country by a substantial 16.9%. The overall changes in welfare are not simply the sum of the two counterfactuals conducted separately, highlighting the interplay between migration restrictions and pollution exposure. For instance, relaxing migration restrictions directly improves welfare by lowering mobility costs and allowing access to better wages; but when combined with lower pollution, the individuals who immigrate also benefit from better air quality.

The numbers in Table 11 make clear that while the policies we study (relocating pollution away from the most productive places and relaxing *hukou*) are sensible on efficiency grounds, they raise welfare by primarily benefiting rich, skilled workers. Poor, unskilled workers are also better off, but their gains are only substantial when we reduce overall mobility costs.

Once again, in Appendix Table E2, we re-examine the overall changes to GDP from our most basic formulation of the model, without externalities. That is, without housing, agglomeration, or pollution responses to changing populations. The results are qualitatively similar to Table 9, with some improvements to GDP being slightly larger in magnitude.

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

Figure 10: 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 equal to the pollution in the median city. The horizontal axis plots the baseline amount of pollution in a city. The vertical axis plots the change in the skilled wage as this baseline pollution is 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-level of pollution the same as before. This means raising pollution levels in cities with low pollution and low skill-biased-capital, and lowering them in polluted, productive cities. In Figure 10 we show how wages would change with this reallocation of pollution, as a way to quantify how much of the wage gap across cities is due to existing patterns of pollution. We use the numbers for Tianjin and Chongqing – two comparable representative provincial-level cities – to illustrate. As pollution is lowered in Tianjin, there is an inflow of skilled workers that lowers the skilled wage. The change in wage is the same as the (endogenously determined) compensating differential. Conversely, skilled 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, ignoring agglomeration or health effects. Once we incorporate the health and agglomeration benefits of reducing pollution, the Tianjin wage does not fall by as much, while the newly imported pollution into Chongqing mitigates the rise in wages there. So a smaller, 14.4% of the gap gets bridged under the new equitable configuration of pollution across cities, once we incorporate the full range of effects of pollution on people’s productivity. Only adding health benefits to the resorting effects (ignoring agglomeration) explains 12.5% of the gap.

9.4 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 $60 \mu\text{g}/\text{m}^3$.” As the logic of our paper would recommend, this plan sensibly targeted China’s three most productive areas. This should induce skilled workers to re-sort into those areas, and raise aggregate productivity.

Table 12: The Productivity Effect of Pollution Regulation

	Change in GDP per Worker in China (%)			
	Overall changes (1)	Health (2)	Relocation (3)	Relocate+Agglom (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 higher tier cities (row 2), and overall migration costs (row 3) by 50%. Column 1 shows the gain to country GDP. Column 2 shows the gain to GDP from the health-productivity channel. Column 3 shows the GDP change from the relocation channel, and Column 4 also accounts for the agglomeration of skilled workers.

As shown in Table 12, we predict this policy – targeted at only a subset of cities – to increase country-level GDP per worker in China 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 re-sorting effects of pollution control leaves some large bills on the sidewalk. Given how sensitive (skilled) Chinese citizens are to pollution, and how responsive they are through their migration choices (as documented in our empirical section), our model clarifies that pairing pollution control efforts with easing *hukou* restrictions or reducing migration costs could produce much larger benefits to society. Table 13 shows that a combined policy of both pollution control and easing mobility would also make the welfare gains more equitable. The 2013 pollution-caps policy raised skilled welfare by 7.2%, but have little effect on the unskilled, who are less sensitive to pollution. However, lowering migration costs as a complementary policy tool would improve welfare for both skilled and unskilled workers.

10 Conclusion

Our analysis highlights the aggregate macroeconomic consequences of an important new pattern of mobility. As economies grow and industrial activity pollutes the environment, workers – especially

Table 13: Welfare Effects of Pollution Regulation

	Skilled Welfare (1)	Unskilled Welfare (2)	Average Welfare (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: In this counterfactual exercise we reduce pollution according to the targets set by the 2013 *Air Pollution Prevention and Control Plan* (row 1). In addition to pollution regulations, we also relax the hukou restriction in higher tier cities (row 2), and migration costs (row 3) by 50%. To capture the distributional implications, we examine how this policy affects wages for different types of workers in Table E3.

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., 2016; 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)? 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. Pollution control coupled with policies easing mobility has the potential to bring about large productivity gains in China.

References

- Adao, R., M. Kolesar, and E. Morales (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics* 134(4), 1949–2010.
- Adhvaryu, A., N. Kala, and A. Nyshadham (2016). Management and shocks to worker productivity. *International Growth Center (IGC)*. Working Paper.
- AFP News (2019, March). Asia’s Pollution Exodus: Firms Struggle to Woo Top Talent. *Agence-France Press News Agency*. <https://news.yahoo.com/asias-pollution-exodus-firms-struggle-woo-top-talent-035837924.html>.
- 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.
- Andreoni, J. and A. Levinson (2001). The Simple Analytics of the Environmental Kuznetz Curve. *Journal of Public Economics* 80(2), 269–286.
- Arceo, E., R. Hanna, and P. Oliva (2016). Does the Effect of Pollution on Infant Mortality Differ between Developing and Developed Countries? Evidence from Mexico City. *The Economic Journal* 126(591), 257–280.
- Arrow, K. J. (1962). Economic Welfare and the Allocation of Resources for Innovation. *NBER working paper c2144*.
- Au, C. and V. Henderson (2006). How Migration Restrictions Limit Agglomeration and Productivity in China. *Journal of Development Economics* 80(2), 350–388.
- Autor, D., D. Dorn, and G. H. Hanson (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review* 103(6), 2121–68.
- Balboni, C. (2021). In Harm’s Way: Infrastructure Investments and the Persistence of Coastal Cities. *Working Paper*.
- Bayer, P., N. Keohane, and C. Timmins (2009). Migration and Hedonic Value: The Case of Air Quality. *Journal of Environmental Resources and Management* 58(1), 1–14.
- Bazzi, S. (2017). Wealth Heterogeneity and the Income Elasticity of Migration. *American Economic Journal: Applied Economics* 9(2), 219–55.
- Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-experimental Shift-share Research Designs. Technical report, National Bureau of Economic Research.
- Brandt, L., C.-T. Hsieh, and X. Zhu (2008). Growth and Structural Transformation in China. *China’s Great Economic Transformation*, 683–728.
- 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.
- Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica* 82(5), 1671–1748.
- Bryan, G. and M. Morten (2019). The Aggregate Productivity Effects 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. G. Zivin, T. Gross, and M. Neidell (2019). The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China. *American Economic Journal: Applied Economics* 11(1), 151–172.
- Che, Y. and L. Zhang (2018). Human Capital, Technology Adoption and Firm Performance: Impacts of China's Higher Education Expansion in the Late 1990s. *Economic Journal* 128(614), 2282–2320.
- Chen, S., P. Oliva, and P. Zhang (2017). The Effect of Air Pollution on Migration: Evidence from China. *NBER Working Papers*.
- Chen, Y. (2020). Pollution Regulations, Local Labor Markets, and Skill Heterogeneity. *Working Paper*.
- Chen, Y., A. Ebenstein, M. Greenstone, and H. Li (2013). Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences of the United States of America* 110(32), 12936–12941.
- Chen, Y., G. Z. Jin, N. Kumar, and G. Shi (2012). Gaming in Air Pollution Data? Lessons from China. *The BE Journal of Economic Analysis & Policy* 12(3), 1–43.
- Clemens, M. A., C. E. Montenegro, and L. Pritchett (2019). The Place Premium: Bounding the Price Equivalent of Migration Barriers. *Review of Economics and Statistics* 101(2), 201–213.
- Combes, P.-P., S. Demurger, S. Li, and J. Wang (2019). Unequal Migration and Urbanisation Gains in China. *Journal of Development Economics* 142(1), 102328.
- Dahl, G. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica* 70(6), 2367–2420.
- Duranton, G. and D. Puga (2005). From Sectoral to Functional Urban Specialisation. *Journal of Urban Economics* 57(2), 343–406.
- Eaton, J. and S. Kortum (2002). Technology, Geography and Trade. *Econometrica* 70(5), 1741–1779.
- Ebenstein, A., M. Fan, M. Greenstone, G. He, P. Yin, and M. Zhou (2015). Growth, Pollution, and Life Expectancy: China from 1991-2012. *American Economic Review* 105(5), 226–31.
- Ebenstein, A., M. Fan, M. Greenstone, G. He, and M. Zhou (2017). New Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. *Proceedings of the National Academy of Sciences* 114(39), 10384–10389.
- Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial Sorting. *Journal of Political Economy* 122(3), 554–620.
- 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.
- Gaubert, C. (2018). Firm Sorting and Agglomeration. *American Economic Review* 108(11), 3117–53.
- 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.
- Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review* 110(8), 2586–2624.
- 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*. Forthcoming.
- Hanlon, W. (2020). Coal Smoke, City Growth, and the Costs of the Industrial Revolution. *The Economic Journal* 130(626), 462–488.
- Harberger, A. (1954). Monopoly and Resource Allocation. *American Economic Review* 44(2), 77–87.
- 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, 18–39.
- 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 (2020). East Side Story: Historic Pollution and Neighborhood Segregation. *Journal of Political Economy*. Forthcoming.
- 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 (2021). Migrants and Firms: Evidence from China. *Working Paper*.
- 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 P. Nilsson (2014). Economic Status, Air Quality, and Child Health: Evidence from Inversion Episodes. *Working Paper*.
- Jia, Barwick, P., S. Li, L. Lin, and E. Zou (2019). From Fog to Smog: The Value of Pollution Information. *National Bureau of Economic Research, WP 26541*.
- Kahn, M. E. and P. Li (2019). The Effect of Pollution and Heat on High Skill Public Sector Worker Productivity in China. *National Bureau of Economic Research, WP 25594*.
- Kapur, M. (2019, November). Many Delhi Residents Want to Escape the Smog. But Where Can They Go? *Quartz India*.
- Khanna, G., K. Shih, A. Weinberger, M. Xu, and M. Yu (2020). Trade Liberalization and Chinese Students in US Higher Education. *Working Paper*.
- Kinnan, C., S.-Y. Wang, and Y. Wang (2018). Access to Migration for Rural Households. *American Economic Journal: Applied Economics* 10(4), 79–119.

- Kone, Z., M. Liu, A. Mattoo, C. Ozden, and S. Sharma (2018). Internal Borders and Migration in India. *Journal of Economic Geography* 18(4), 729–759.
- Lagakos, D., M. Mobarak, and M. Waugh (2019). The Welfare Effects of Encouraging Rural-urban Migration. *Working Paper*.
- Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labor. *The Manchester School* 22(2), 139–191. May.
- Lustgarten, A. (2020, September). How Climate Migration Will Reshape America. *New York Times*.
- 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.
- Munshi, K. and M. Rosenzweig (2016). Networks and Misallocation: Insurance, Migration, and the Rural-urban Wage Gap. *American Economic Review* 106(1), 46–98.
- 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.
- 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 Meteorology and Climatology* 53(4), 840–858.
- Restuccia, D. and R. Rogerson (2013). Misallocation and Productivity. *Review of Economic Dynamics* 16(1), 1–10.
- Restuccia, D. and R. Rogerson (2017). The Causes and Costs of Misallocation. *Journal of Economic Perspectives* 31(3), 151–74.
- 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 (2020, September). Brain Drain: The Impact of Air Pollution on Firm Performance. *Available on SSRN*. <https://ssrn.com/abstract=3490344> or <http://dx.doi.org/10.2139/ssrn.3490344>.
- Young, A. (2014). Structural Transformation, the Mismeasurement of Productivity Growth, and the Cost Disease of Services. *American Economic Review* 104(11), 3635–67.
- Yu, B., W.-S. Lee, and S. Rafiq (2020). Air Pollution Quotas and the Dynamics of Internal Skilled Migration in Chinese Cities. *Working Paper*.
- 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.

Appendix

Table of Contents

A	Robustness of Empirical Specifications	A2
A.1	The Different Instruments and their Sources of Variation	A2
A.2	The Huai River Regression Discontinuity	A11
A.3	Alternative Model Specifications, Controls and Samples	A13
B	Air Pollution Data Disclosure in China	A21
C	Additional Tables and Figures	A22
D	Additional Model Derivations	A28
D.1	Deriving Labor Supply and Welfare	A28
D.2	Elasticity of Capital, and Modelling Skill-biased Capital	A29
D.3	A Note on Existence and Uniqueness of the Equilibrium	A30
E	Additional Counterfactual Results	A31

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 instrumental variables, alternative model specifications and data sources.

A.1 The Different Instruments and their Sources of Variation

First, as we describe in the main text, we explore a few different sources of underlying variation. We employ two different instrumental variables strategies discussed in recent work on China to address the endogeneity of air pollution. Specifically, we study the variation underlying the instrument based on wind direction and distant coal-fired power plants (Freeman, Liang, Song, and Timmins, 2019), and the variation in air quality driven by the number and strength of thermal inversions (Chen, Oliva, and Zhang, 2017).

Then, we explore threats to identification for our two main instruments. We test concerns of the endogenous placements of power plants, whereby policy makers may use the same function – the simultaneous interaction between wind directions, distance to 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, and find a similar empirical pattern. We may still think that *newly* built plants are endogenously placed. Yet, our results are robust to relying on old power plants, and to make it even more conservative the 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.

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*** (0.00211)	0.0113*** (0.00235)	0.00956*** (0.00239)						
Number of inversions				0.00222*** (0.000350)	0.00159*** (0.000329)	0.00201*** (0.000293)			
Strength of inversions							0.000730*** (0.000214)	0.000448** (0.000185)	0.000725*** (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*** (3.94e-05)	0.0136*** (3.44e-05)	0.0130*** (3.38e-05)						
Number of inversions				0.00212*** (6.24e-06)	0.00184*** (6.08e-06)	0.00227*** (5.30e-06)			
Strength of inversions							0.000822*** (3.85e-06)	0.000670*** (3.77e-06)	0.000944*** (3.29e-06)
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. 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.

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

Table A2: Stacked Regressions: With High-skill Interaction

	Leave hukou city indicator		
	(1) Wind IV	(2) Thermal Num	(3) Thermal Strength
Log(PM2.5)	0.0751* (0.0388)	0.0845*** (0.0203)	0.0704*** (0.0230)
Log(PM2.5)*High-skill	0.0622** (0.0251)	0.0810** (0.0376)	0.110** (0.0528)
Observations	761,548	761,548	761,548
R-squared	0.025	0.033	0.034
City Controls	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Weather Controls	No	Yes	Yes

Notes: Individual level regressions across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant, and using the number of thermal inversions and the strength of thermal inversions. We also control for a high-skilled indicator. 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.

In 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 Table A3, we study how variation in PM2.5 from our various instruments affects migration. Panel A shows the results without weather controls. This replicates the results in the main text for the coal-fired plants IV in the first three columns. We include distances to three large seaports to account for the spatial distribution of economic development in China. In the next three columns, we employ the annual occurrence of thermal inversions as the instrumental variable. The results show that the implications of air pollution on emigration are more pronounced for high-skilled workers in comparison with those with lower skills. In the last three columns, we leverage the variation coming from the annual strength of thermal inversions. The results remain similar. Since thermal inversions may be affected by local weather conditions, we further control for a set of weather characteristics. Panel B of Table A3 shows these patterns with weather controls. We find a similar empirical pattern: high-skilled workers are more likely to escape from polluted cities to avoid the adverse effects of pollution, compared to their lower educated counterparts.

Table A3: Pollution and Out-migration

Panel A: No weather	Dependent variable: Leave hukou city indicator								
	Wind+Coal IV			Number of inversions			Strength of inversions		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)	Full sample (7)	Low edu (8)	High edu (9)
Log (PM2.5)	0.0772** (0.0389)	0.0609 (0.0423)	0.140*** (0.0382)	0.113*** (0.0274)	0.108*** (0.0291)	0.154*** (0.0358)	0.0970*** (0.0341)	0.0920** (0.0358)	0.168*** (0.0553)
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	No	No	No	No	No	No	No	No	No
F-test of IVs	52.53	46.52	41.54	52.33	50.27	39.10	19.82	20.24	11.58

Panel B: Weather controls	Dependent variable: Leave hukou city indicator								
	Wind+Coal IV			Number of inversions			Strength of inversions		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)	Full sample (7)	Low edu (8)	High edu (9)
Log (PM2.5)	0.0890*** (0.0343)	0.0763** (0.0368)	0.155*** (0.0359)	0.0906*** (0.0202)	0.0871*** (0.0211)	0.112*** (0.0233)	0.0779*** (0.0228)	0.0740*** (0.0238)	0.107*** (0.0279)
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	47.20	41.91	38.24	102.5	97.33	84.83	51.09	49.60	40.54

Notes: Individual level regressions across 332 cities. Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant, 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.

Finally, Table A4 shows combinations of the Wind IV and thermal inversions instruments, and the corresponding over-identification tests. Across specifications, we find similar results.

While our earlier results examine the out-migration decision of individuals, we further examine the association between air pollution and destination choices. We regress city-level in-migration rate on local air pollution concentration. The results are in Table A5. We use our instruments to deal with the endogeneity of air quality in destination cities. We find that even for in-migration decisions, the response of high-skilled workers is greater than that of low-skilled workers. In other words, high-skilled workers are more likely to move to cities with clean air when they make location choices. Severe air pollution can not only result in the outflow of high-skilled workers but also reduce their inflow.

A.1.2 Endogeneity Concerns over Instrumental Strategies

In this section we test the identification assumptions underlying our main instruments. 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. For example, those plants may be systematically built near poorer, less influential cities. Thus, there is a concern that the instrument may be correlated with unobservable characteristics of nearby cities. In Table A6, we exclude any plants built within 200km of a given city (first two columns), and then within 400km of the city (last two columns). We still

Table A4: Combined Instruments

	Dependent variable: Leave hukou city indicator					
	Wind and Coal +Number of inversions			Wind and Coal +Strength of inversions		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (PM2.5)	0.0892*** (0.0189)	0.0836*** (0.0197)	0.124*** (0.0228)	0.0795*** (0.0206)	0.0722*** (0.0213)	0.129*** (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. 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.

restrict the radius to be of 400km widths. Our results are similar to before, with a slight increase in precision.

Even though policy makers may not have modeled the polluting potential of plants built in the past in the manner that we do, we may expect that newly built plants are subject to more scrutiny as the conversation about air quality in China has recently escalated. In Table A7 we exclude newly built plants in the construction of IV, and instead include cities with newly built ones in the ‘control’ group. Although this empirical strategy is more conservative, we find similar patterns and magnitudes.

We may also expect that coal-fired plants are more likely to be located in coal producing areas. In as far as coal producing areas may also influence the underlying industrial structure, this may raise concerns about other unobservable associations with migration rates. Shanxi is the largest coal producing province in China, we thus exclude Shanxi in our estimation. As reported in in Table A8, the results are slightly more precisely estimated.

Our wind-direction and coal-plants 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 A9 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 at power plant. 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 A10, we present the number of large-scale thermal plants located up wind of five largest metropolitan areas in 2014 along with their total coal consumption. Beijing and Tianjin are among the most populated and politically important cities in Northern China, but have high levels of pollution.

Table A5: 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 (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)	Full sample (7)	Low edu (8)	High edu (9)
Log (PM2.5)	-1.856*** (0.529)	-1.853*** (0.520)	-2.259*** (0.716)	-1.204** (0.500)	-1.137** (0.475)	-1.645** (0.677)	-1.404** (0.643)	-1.275** (0.606)	-2.152** (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 (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)	Full sample (7)	Low edu (8)	High edu (9)
Log (PM2.5)	-1.833*** (0.529)	-1.825*** (0.513)	-2.266*** (0.728)	-1.204*** (0.373)	-1.181*** (0.357)	-1.451*** (0.503)	-1.359*** (0.421)	-1.301*** (0.401)	-1.794*** (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. Robust standard errors are reported in parentheses. 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.

There are 58 large scale coal-fired plants located upwind of Beijing, and 59 large scale thermal power plants located upwind of Tianjin. The total consumption of them is as large as 2372 million tons and 2123 million tons respectively. We also calculate the ratio of the upwind large thermal plants to the total number of large thermal plants. As presented in column 2, this ratio 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 in a manner that simultaneously incorporates the interactions between wind direction, distance and coal consumption, we explore whether baseline city features predict newly built plants. In Table A11, 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 and major provincial capitals (Table A18).

In Table A12, we construct placebo instruments, artificially changing the wind direction angle by 90 degrees, and then 180 degrees. The first two columns report the first stage results. As the angle is increased, the falsified instrument is less likely to predict PM2.5 or emigration.

Table A6: Different Distance Bins for Selection of Plants

	Dependent variable: Leave hukou city indicator					
	Distance 200-600km			Distance 400-800km		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (PM2.5)	0.0947*** (0.0365)	0.0798** (0.0396)	0.163*** (0.0383)	0.0973*** (0.0304)	0.0834*** (0.0319)	0.168*** (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. 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.

Table A7: Excluding Newly Built Power Plants

	Dependent variable: Leave hukou city indicator					
	Plants > 5 yrs ago			Plants > 10 yrs ago		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (PM2.5)	0.0775** (0.0387)	0.0611 (0.0421)	0.138*** (0.0379)	0.0916** (0.0415)	0.0761* (0.0451)	0.146*** (0.0412)
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	50.37	44.39	40.57	38.94	33.08	34.42

	Dependent variable: Leave hukou city indicator					
	Plants > 15 yrs ago			Plants > 20 yrs ago		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (PM2.5)	0.0920** (0.0397)	0.0762* (0.0432)	0.150*** (0.0401)	0.111*** (0.0428)	0.0992** (0.0469)	0.156*** (0.0398)
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	42.63	35.28	42.99	37.18	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. 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 A8: Without Coal Producing Regions

Dependent variable: Leave hukou city indicator			
	Full sample	Low edu	High edu
	(1)	(2)	(3)
Log (PM2.5)	0.0921*** (0.0349)	0.0773** (0.0375)	0.148*** (0.0345)
Observations	722,306	609,605	112,701
City Controls	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
F-test of IVs	84.82	77.61	68.94

Notes: Individual level regressions across 321 cities (we drop coal producing regions, like all cities in Shanxi province). Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include the log distance to Shanghai seaport, to Tianjin seaport and to Shenzhen seaport and to the nearest big city. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A9: 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** (0.0392)	0.0785* (0.0424)	0.163*** (0.0392)	0.0905** (0.0358)	0.0765** (0.0385)	0.144*** (0.0350)	0.0360 (0.0343)	0.0199 (0.0378)	0.0911*** (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. 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.

Table A10: 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, Liang, Song, and Timmins \(2019\)](#), 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 A11: 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 (0.0309)	-0.0132 (0.0279)	0.412 (0.489)	0.226 (0.440)
Baseline GDP per capita	-0.0151 (0.0302)	-0.0231 (0.0289)	-0.132 (0.526)	-0.414 (0.522)
Industrial Elec cons	0.0225 (0.0224)		0.210 (0.347)	
Baseline Elec cons		0.0255 (0.0178)		0.414 (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 A12: Placebo Wind Directions

	Log (PM2.5)		Leave hukou city indicator			
	(1)	(2)	Coal IV Placebo (+90 degrees) Low edu (3)	Coal IV Placebo (+90 degrees) High edu (4)	Coal IV Placebo (+180 degrees) Low edu (5)	Coal IV Placebo (+180 degrees) High edu (6)
Log (PM2.5)			-0.0708 (0.0594)	0.0235 (0.0518)	-0.159 (0.157)	-0.0281 (0.0760)
Coal IV Placebo (wind direction+90 degrees)	0.00957** (0.00478)					
Coal IV Placebo (wind direction+180 degrees)		0.00134 (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. 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.

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., 2017; Hicks et al., 2015; Jans et al., 2014), 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 A13. 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.

Table A13: Lagged Pollution and Thermal Inversions

	Dependent variable: Strength of inversions		
	(1)	(2)	(3)
Lagged Log(PM2.5)	-11.68 (12.88)	-11.79 (12.77)	-11.38 (12.80)
Lagged number of inversions			0.241 (0.181)
Lagged strength of inversions		0.0332 (0.0330)	-0.0373 (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

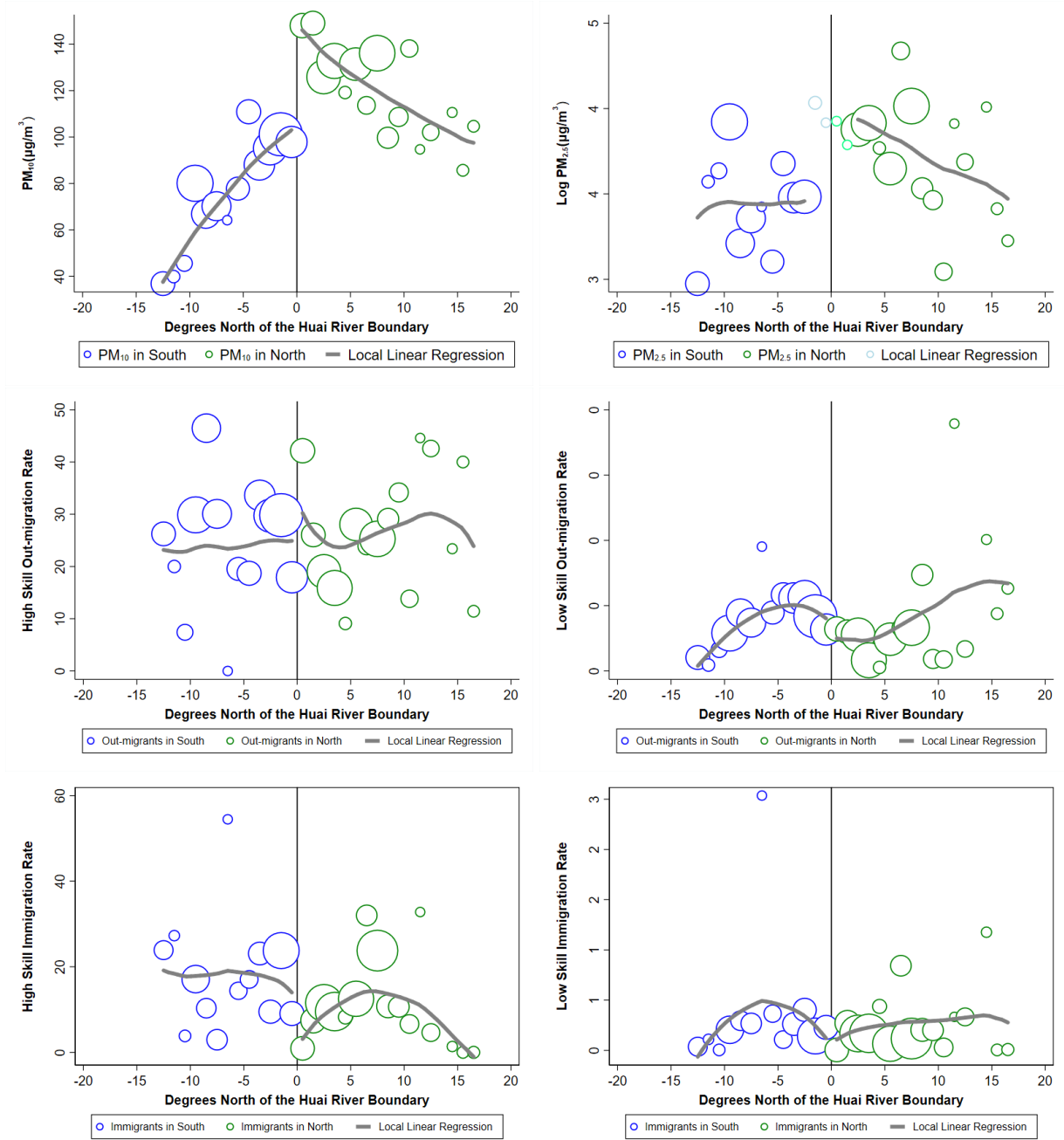
Notes: Standard errors clustered at city level are reported in parentheses. City level regressions for 332 cities over 11 years (2005 to 2015). Specifications include city and year fixed effects.

A.2 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 which releases a large amount of pollutants. [Chen et al. \(2013\)](#) examine the effects of this policy on life expectancy using a regression discontinuity design where they compare cities just north of the river to those just south of it.

Here, we leverage the same empirical setup to examine migration decisions. Figure [A1](#) shows the RD graphs where the horizontal axis represents the distance between the city and the Huai river. In our top row we show the discontinuity in PM10 and PM2.5 levels. In the middle row, we look at out-migration decisions, and fail to find any meaningful changes in out-migration rates. In the bottom row, we look at in-migration rates, and find that only for the high-skill workforce, there is less in-migration in cities that have more pollution. This 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.3 Alternative Model Specifications, Controls and Samples

In this section we examine different model specifications, sample restrictions and control variables. Once again, our aim is to test the robustness of our estimates to various empirical concerns.

First, we employ an individual-level longitudinal panel data to estimate the association between air quality and workers' spatial sorting. The longitudinal data allow us to track individuals over time and control for individual-level unobservables. Importantly, we use an alternative definition for migration status regardless of *hukou* location, and show results similar to before.

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

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

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

A.3.1 Individual Longitudinal Panel and Alternative Definition of Migration

We employ an individual-level longitudinal panel and a different definition of migration status to explore the spatial sorting of Chinese workers. We use an alternative data source, the China Labor-force Dynamic Survey (CLDS), which is a national social survey targeted at labor force dynamics in China. CLDS 2016 asks a retrospective history of locations for individuals. We use this retrospective location history to 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* location or not. The strengths of the individual-level panel lie in that it allows us to account for individual-specific unobservables and track those who have moved multiple times and who have moved and returned home. We combine the longitudinal data with annual information on PM2.5 at the city level.

Since data on large-scale coal-fired plants are not available for each year during the sample period, we employ the instrument of the number of thermal inversions to address the endogeneity of air pollution. Table A14 presents the IV estimation of the relationship between air pollution and the out-migration tendency of individuals, controlling for year- and individual- fixed effects. Including individual-specific 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 systematically correlated with migration decisions. The results show that once again, the response of high-skill migration to air pollution is a lot stronger than that of low-skill workers. A 10% increase in PM2.5 raises out-migration rates by 2.19 percentage points for high-skilled workers.

In our main text, migrants are defined as those who are away from their *hukou* city. Under this definition, we may miss those who move to a different city and obtain local *hukou* in the city. Thus, we may understate high-skilled workers' migration response to pollution, 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* 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 A14 are larger in magnitude than our baseline estimates, perhaps as here we measure both *hukou* and non-*hukou* migration.

We control for weather amenities in the next three columns. The coefficient estimates are quantitatively and qualitatively similar. In China, migrant workers are more likely to move from the inland to coastal regions to gain access to better economic opportunities. As a result, the vast majority of

Table A14: Individual Longitudinal Panel with Individual Fixed Effects

	Dependent variable: Leave city location indicator								
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)	Full sample (7)	Low edu (8)	High edu (9)
Log (PM2.5)	0.115** (0.0558)	0.0812 (0.0542)	0.219** (0.100)	0.0960* (0.0540)	0.0644 (0.0525)	0.193** (0.0964)	0.0774 (0.0539)	0.0385 (0.0538)	0.206** (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 city level are reported in parentheses. 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.

migrants are concentrated in large coastal cities. 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 do not affect our results meaningfully.

A.3.2 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 large-scaled 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 A15, we use specifications where PM2.5 are averaged over 5, 10, 15 and 18 years, respectively, leading up to the migration decision. We find that the longer the time period of PM2.5 exposure, the larger the response in small increments. As such, pollution exposure in a short time frame ha 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 skilled than the unskilled.

In Table A16, we further add cumulative weather conditions that may confound the association between the thermal inversions IV and migration decisions. We consistently find stronger effects on out-migration for the higher educated group of workers.

A.3.3 Alternative Samples, Controls and Alternative Measure of Air Quality

In this section, we use different samples to explore the relationship between pollution and out-migration. We then use an alternative measure of air quality as your main independent variable.

First, we look at different ways to reformulate the estimation sample. In Table A17, 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 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.

Table A15: PM2.5 Measured over Different Time Intervals

	Dependent variable: Leave hukou city indicator					
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (Mean PM2.5: 1998-2015)	0.138*** (0.0346)	0.132*** (0.0364)	0.202*** (0.0493)			
Log (Mean PM2.5: 2001-2015)				0.134*** (0.0334)	0.128*** (0.0352)	0.195*** (0.0473)
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	37.87	35.91	27.09	40.25	38.20	28.65

	Dependent variable: Leave hukou city indicator					
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (Mean PM2.5: 2006-2015)	0.126*** (0.0310)	0.120*** (0.0326)	0.185*** (0.0445)			
Log (Mean PM2.5: 2011-2015)				0.117*** (0.0290)	0.111*** (0.0307)	0.173*** (0.0415)
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	41.72	39.68	30.12	42.11	40.08	31.10

Notes: Individual level regressions across 332 cities. Instrumental variables using the number of thermal inversions averaged over different time intervals. Standard errors clustered at the hukou city level are reported in parentheses. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

We next examine whether big cities, high polluting cities, or major province capitals are driving our results. We do this by excluding such cities one at a time in Table A18. This may help allay concerns about the influence of major cities or province capitals in pollution policy, placement of plants, or being outliers in terms of pollutants and/or skilled jobs.

We further conduct heterogeneity analysis. In Table A19, we split up the sample into two age groups. We see that the implications of pollution on emigration are stronger for younger workers. There are two reasons for the differential migration pattern by age. First, young people may have lower mobility costs than their older counterparts. Second, they may have better knowledge about the adverse effects of air pollution. In Panel B, we examine the effects by marital status. Married individuals are more responsive, perhaps as they may have young children that may be affected by poor air quality.

In Table A20 we split up the sample by rural and urban origin locations. We find that the elasticity for high-skill out-migration is larger in rural areas than it is in urban areas.

Finally, we turn our attention to an alternative measure of local air quality. As sources of pollution affect not just 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, PM_{2.5}, O₃ and CO. Furthermore, the AQI is officially reported and widely disseminated, and certain individuals may be more likely to respond to it. As shown in Table A21, our empirical pattern are robust to using the AQI as our

Table A16: PM2.5 Measured over Different Time Intervals: Controlling for Weather Conditions

	Dependent variable: Leave hukou city indicator					
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (Mean PM2.5: 1998-2015)	0.113*** (0.0236)	0.104*** (0.0235)	0.173*** (0.0394)			
Log (Mean PM2.5: 2001-2015)				0.107*** (0.0230)	0.0979*** (0.0231)	0.168*** (0.0387)
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	65.36	63.86	44.57	68.08	66.49	45.59
	Dependent variable: Leave hukou city indicator					
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (Mean PM2.5: 2006-2015)	0.122*** (0.0264)	0.114*** (0.0273)	0.167*** (0.0359)			
Log (Mean PM2.5: 2011-2015)				0.0913*** (0.0220)	0.0850*** (0.0231)	0.129*** (0.0297)
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	80.66	78.19	57.01	91.53	87.48	71.38

Notes: Individual level regressions across 332 cities. Instrumental variables using the number of thermal inversions averaged over different time intervals. Standard errors clustered at the hukou city level are reported in parentheses. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator. Weather controls include temperature, humidity, sunshine duration and wind speed averaged over different time intervals.

independent variable of interest.

A.3.4 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 A22 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 A24 we test the validity of the wind IV to other economic and industrial controls. In the first few columns of Table A24 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.

In the next three columns of Table A24 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.

Table A17: Disaggregated Education Levels

	Dependent variable: Leave hukou city indicator		
	High school or below (1)	Some college (2)	College or above (3)
Log (PM2.5)	0.0609 (0.0423)	0.116*** (0.0401)	0.176*** (0.0409)
Observations	643,124	64,598	53,826
City Controls	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
F-test of IVs	46.52	44.18	34.59

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

Table A18: Without Big Cities, High Polluters and Major Province Capitals

	Exclude Beijing			Exclude Tianjin			Exclude Shijiazhuang		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)	Full sample (7)	Low edu (8)	High edu (9)
Log (PM2.5)	0.0761* (0.0389)	0.0609 (0.0424)	0.132*** (0.0369)	0.142** (0.0556)	0.129** (0.0619)	0.180*** (0.0487)	0.0783** (0.0393)	0.0623 (0.0426)	0.140*** (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

	Exclude Shenyang			Exclude Zhengzhou			Exclude Wuhan		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)	Full sample (7)	Low edu (8)	High edu (9)
Log (PM2.5)	0.0766** (0.0387)	0.0605 (0.0421)	0.138*** (0.0377)	0.0816** (0.0397)	0.0641 (0.0431)	0.155*** (0.0377)	0.0775** (0.0392)	0.0612 (0.0425)	0.141*** (0.0390)
Observations	758,589	641,096	117,493	754,535	637,980	116,555	758,065	640,833	117,232
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	52.22	46.37	40.90

Notes: Individual level regressions across 331 cities (we exclude one big city at a time). Standard errors clustered at the hukou city level are reported in parentheses. Instrumental variables specification using the interaction between wind direction, distance to coal plant, and coal consumption at power plant. City controls include distance to Shanghai, Tianjin and Shenzhen seaports. Demographics include age, age-squared, gender, marital status, and an urban hukou indicator.

Table A19: By Age Groups

	Dependent variable: Leave hukou city indicator					
	Age<35			Age>=35		
	Full Sample (1)	Low edu (2)	High edu (3)	Full Sample (4)	Low edu (5)	High edu (6)
Log (PM2.5)	0.114** (0.0496)	0.0640 (0.0627)	0.193*** (0.0504)	0.0604* (0.0342)	0.0572 (0.0368)	0.0684*** (0.0227)
Observations	270,601	199,958	70,643	490,947	443,166	47,781
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	48.21	33.65	40.70	54.11	49.93	40.62

	Age<35 & Unmarried			Age<35 & Married		
	Full Sample (1)	Low edu (2)	High edu (3)	Full Sample (4)	Low edu (5)	High edu (6)
	Log(PM2.5)	0.0666** (0.0336)	0.00214 (0.0435)	0.140*** (0.0370)	0.144** (0.0605)	0.0922 (0.0737)
Observations	104,009	66,956	37,053	166,592	133,002	33,590
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
F-Test of IVs	43.86	30.24	42.93	49.21	34.24	36.70

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

Table A20: Urban vs Rural Outmigration

	Dependent variable: Leave hukou city indicator					
	Urban hukou			Rural hukou		
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
Log (PM2.5)	0.0589** (0.0283)	0.0386 (0.0299)	0.110*** (0.0313)	0.106* (0.0641)	0.0899 (0.0647)	0.342*** (0.0932)
Observations	401,351	302,344	99,007	360,197	340,780	19,417
City Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
F-test of IVs	58.20	58.24	38.10	30.70	28.98	35.71

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

Table A21: Air Quality Index

Dependent variable: Leave hukou city indicator			
	Full sample	Low edu	High edu
	(1)	(2)	(3)
Log (AQI)	0.0795 (0.0659)	0.0581 (0.0682)	0.176** (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. 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.

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 A24 Panel A. The inclusion of these industrial emissions does little to affect the impacts of PM_{2.5} 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 plants may be built near cities that require more electricity. Even though those plants supply electricity to vast areas of China including many remote provinces (Freeman, Liang, Song, and Timmins, 2019), we examine this concern by controlling for city-level electricity consumption. We include total (industrial, commercial and residential) electricity consumption in the next three columns of Table A24 Panel B, and add industrial electricity consumption in the last three columns. Accounting for city-level electricity consumption barely changes our empirical patterns.

Table A22: 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)	0.112*** (0.0436)	0.110** (0.0461)	0.155*** (0.0474)	0.0837* (0.0508)	0.0787 (0.0534)	0.151** (0.0589)
Observations	671,202	562,938	108,264	671,202	562,938	108,264
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	36.76	34.06	34.77	13.85	13.15	13.20

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. 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.

Table A23: 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** (0.121)	0.314** (0.147)	0.204* (0.119)
Log(L _{hd} /L _{ud})	0.423*** (0.0583)	0.407*** (0.0690)	0.456*** (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. 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.

Table A24: Additional Controls: Skills, Economic and Industrial Emissions

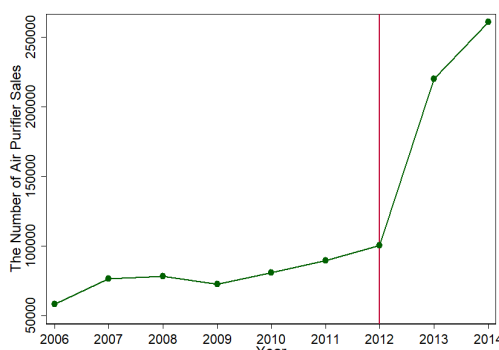
Panel A		Dependent variable: Leave hukou city indicator								
Additional Controls:	Baseline skill controls			Baseline economy controls			Industrial emmissions controls			
	Full Sample (1)	Low edu (2)	High edu (3)	Full Sample (4)	Low edu (5)	High edu (6)	Full Sample (7)	Low edu (8)	High edu (9)	
Log (PM2.5)	0.0712 (0.0433)	0.0592 (0.0466)	0.124*** (0.0427)	0.0403 (0.0422)	0.0251 (0.0454)	0.0984** (0.0444)	0.0539 (0.0435)	0.0396 (0.0470)	0.119*** (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 (1)	Low edu (2)	High edu (3)	Full Sample (4)	Low edu (5)	High edu (6)	Full Sample (7)	Low edu (8)	High edu (9)	
Log (PM2.5)	0.0756** (0.0343)	0.0677* (0.0362)	0.109*** (0.0335)	0.0653* (0.0369)	0.0587 (0.0393)	0.0959*** (0.0346)	0.0643* (0.0358)	0.0574 (0.0380)	0.0972*** (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. 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.

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. Since 2000, the Chinese government released air pollution data in two waves: First, in 2000, the government started publishing air quality data, including an Air Pollution Index (API) and PM10, but only for 42 cities. Although fine particles (i.e. PM2.5) are more hazardous than larger particles (i.e. PM10) with respect to mortality, cardiovascular and respiratory endpoints, PM2.5 was not included in the API. The number of cities in which API and PM10 were available increased to 120 in 2012.

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



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

Second, 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.

Table B1: PM2.5 Data Disclosure and Outmigration

Dependent variable:	Leave city location indicator					
	Full sample (1)	Low edu (2)	High edu (3)	Full sample (4)	Low edu (5)	High edu (6)
$\text{Log}(\text{PM2.5})_{ot} \times \text{Data disclosed indicator}_{ot}$	0.000902 (0.00147)	-0.000585 (0.00139)	0.0106** (0.00495)	0.000997 (0.00143)	-0.000482 (0.00130)	0.0105** (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. 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.

The disclosure of pollution information had an important effect on household avoidance behavior. Figure B1 shows how the sales of indoor air filtration increased sharply 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.

C Additional Tables and Figures

Table C1: Summary Statistics

Variable name	Description	Mean	Std. dev
Out migration	Indicator=1 if the person left his/her hukou city for more than 6 month, =0 otherwise	0.10	0.30
Female	Indicator=1 if the person is female	0.50	0.50
Age		39.87	10.99
Urban hukou	Indicator =1 if the person holds urban hukou, =0 otherwise	0.53	0.50
Married	Indicator =1 if the person is married	0.82	0.39
High-skill	Indicator=1 if the highest degree is some college or above, =0 otherwise	0.16	0.36
Pollution Levels	Annual PM2.5 concentration	38.95	16.67
GDP per capita		81002.09	61579.87
Distance to Tianjin seaport		1223.47	677.02
Distance to Shanghai seaport		1289.79	758.51
Distance to Shenzhen seaport		1470.92	835.869
Temperature	Annual mean temperature	141.79	54.60
Wind speed	Annual mean wind speed	45.47	336.07
Humidity	Annual mean humidity	66.98	11.91
Sunshine duration	Annual total sunshine duration	20305.59	5775.47

Note: Table shows summary statistics for most outcomes, control variables, and determinants of the different samples. In most regressions we control for distance to the three largest seaports: Tianjin seaport, Shanghai seaport, and Shenzhen seaport. These seaports are located at the three major economic circles of China: Beijing-Tianjin-Hebei Metropolitan Region, The Yangtze River Delta, the Pearl River Delta.

Table C2: Examples of 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. Authors' calculations using the official government websites of the Beijing, Shanghai, Guangzhou and Shenzhen. '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 C3: Preferences for Environmental Issues by Education Levels

Dependent Variable: The Environmental issue in China is Terrible	
High school	0.133*** (0.0122)
Some college or above	0.176*** (0.0145)
p-values	0.002
t-values	-3.17
The average value for below high school	0.549
Demographics	Y
City FE	Y
N	24538
adj. R2	0.115

Note: Individual level regressions from 113 cities. Demographic controls include age, age-squared, gender, marital status and an indicator for urban *hukou*. The data source is the China Household Panel Survey 2016 (CFPS). In the CFPS 2016, there is a survey question: In your opinion, how terrible the environment issue is in China. (0=totally not terrible; 2,...,10=very terrible). Based on this question, we define environmental attitude dummy: D=1, if the answer is 6-10; =0, if the answer is 0-5. P-value: the p-values of test of Some college or above=High school; t-value: t-values of test of Some college or above=High school. Standard errors clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C4: Concerns and Actions Taken on Environmental Issues

Panel A: Concerns on Environmental Issues			
	Discuss environmental issues (1)	Donation for environment protection (2)	Concern over environmental issue (3)
High school	0.134*** (0.0171)	0.0585*** (0.0115)	0.120*** (0.0154)
Some college or above	0.231*** (0.0188)	0.151*** (0.0153)	0.174*** (0.0189)
P-value	0.000	0.000	0.001
t-value	-6.01	-5.71	-3.56
Baseline average	0.392	0.106	0.402
City FE	Y	Y	Y
Demographics	Y	Y	Y
N	11147	11147	11147
adj. R2	0.190	0.171	0.192

Panel B: Actions Taken on Environmental Issues			
	Appeal on Environmental issue (1)	Government environmental activity (2)	Non-government environmental activity (3)
High school	0.0255*** (0.00845)	0.119*** (0.0140)	0.0690*** (0.0118)
Some college or above	0.0574*** (0.0120)	0.246*** (0.0169)	0.156*** (0.0144)
P-value	0.010	0.000	0.000
t-value	-2.62	-8.21	-6.31
Baseline average	0.0597	0.135	0.100
City FE	Y	Y	Y
Demographics	Y	Y	Y
N	11147	11147	11147
adj. R2	0.176	0.186	0.194

Note: Individual level regressions from 127 cities. Demographic controls include age, age-squared, gender, marital status and an indicator for urban *hukou*. The data source is the Chinese General Social Survey (CGSS). In the CGSS, there is a survey question: whether you participate in the following activity. 1=never, 2=occasionally; 3=often. We define an indicator: D=1 if the answer=2,3; D=0 if the answer=1. P-value: the p-values of test of Some college or above=High school; t-value: t-values of test of Some college or above=High school. Standard errors clustered at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C5: 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*** (0.0161)	0.162*** (0.0153)		
WID IV			0.00222*** (0.000123)	0.00387*** (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)
Δ College graduates ₂₀₀₁₋₅	0.400*** (0.0386)	0.395*** (0.0390)		
Δ College graduates ₂₀₀₁₋₅ /Baseline $L_{it,2001}$			0.835*** (0.130)	0.127 (0.0915)
Predicted migration flow			0.00158 (0.00148)	0.00425*** (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

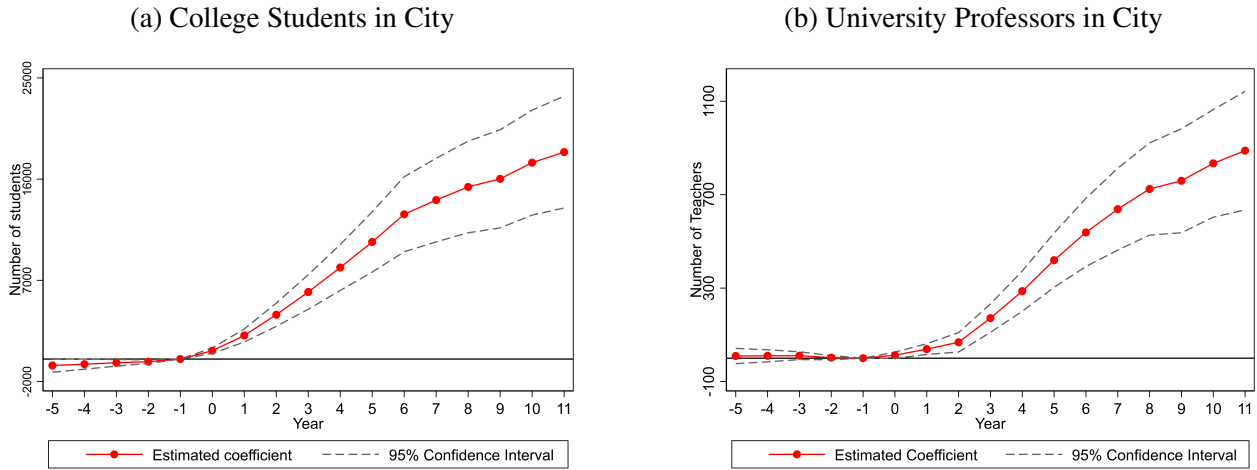
Notes: We control for number of inversions in Column 1 and 3, and strength of inversions in Column 2 and 4. All regressions weighted by the population in 2000. Panel A: We control for hukou index, the interaction of hukou index to migration status indicator, the inverse hyperbolic sine of distance from 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

Table C6: Correlation between College-IV and Pollution Instruments

	(1)	(2)	(3)	(4)	(5)	(6)
	Wind IV	Wind IV	Thermal Num	Thermal Num	Thermal Strength	Thermal Strength
College Expansion IV	0.699 (0.733)	-0.795 (0.690)	0.000893 (0.000626)	-7.48e-05 (0.000598)	-0.00154 (0.00120)	0.000407 (0.00117)
Observations	121	121	121	121	121	121
R-squared	0.007	0.327	0.016	0.333	0.012	0.382
City Control	No	Yes	No	Yes	No	Yes
Weather Control	No	No	No	Yes	No	Yes

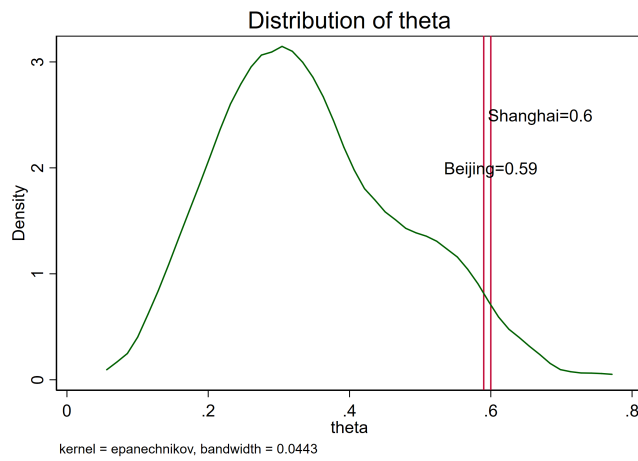
Notes: Tables test for the correlation between the college-expansion IV and the primary instruments used for changes in pollution. City level regressions. City controls include the log distance to Shanghai seaport, to Tianjin seaport and to Shenzhen seaport. Weather controls include temperature, wind speed, sunshine duration and humidity.

Figure C1: Event Study of the Number of College Students by Baseline Propensity to Expand



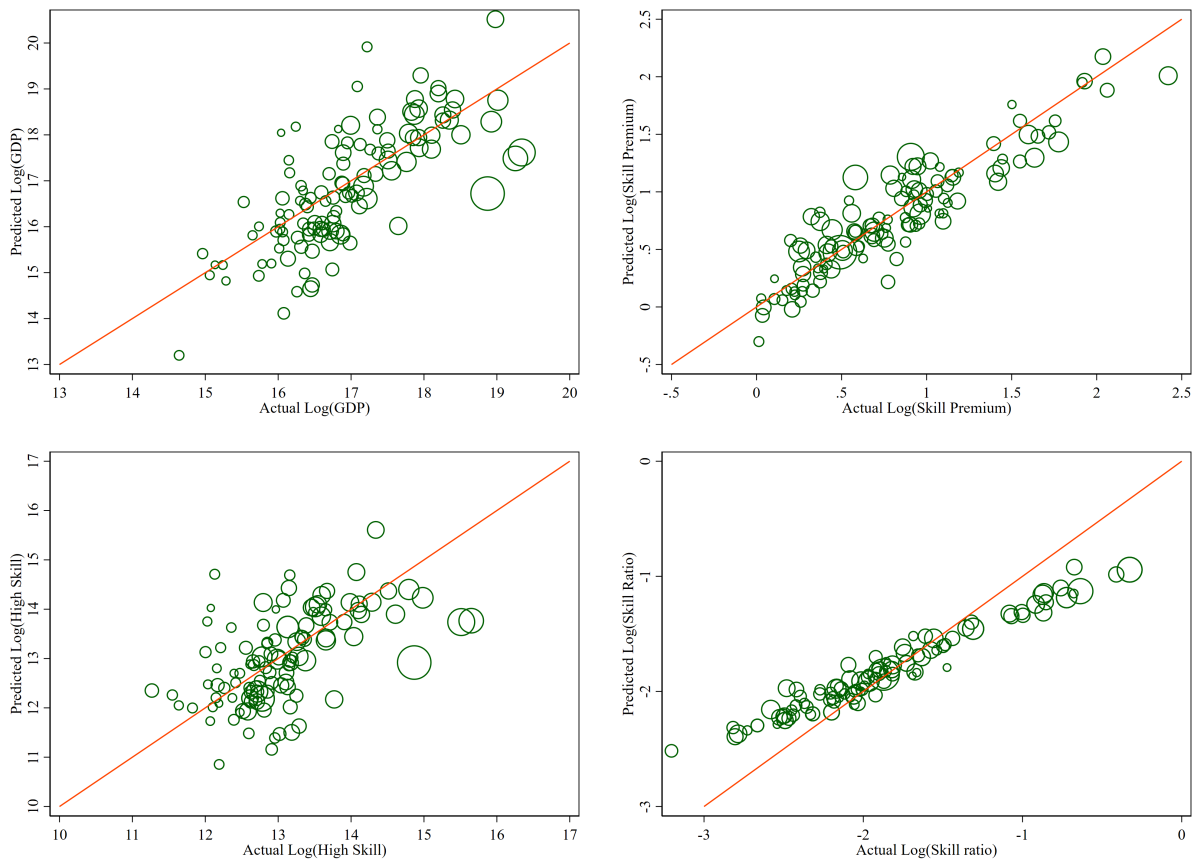
Notes: We test for pre-trends and dynamics of the college expansion policy in an event study framework. City-by-year level regressions from 1994 to 2010. Outcome is number of new college students, relative to one year preceding the expansion (1998). In Figure C1a, we run the regression: $Students_{dt} = \beta_0 + \sum_{t=1994}^{2010} \beta_t (1_{t-1998} \times Treat_{d,1990}) + \gamma_i + \delta_c + \epsilon_{dt}$. In Figure C1b 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 C2: 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 C3: Model Fit and Validation in 2015



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

D Additional Model Derivations

D.1 Deriving Labor Supply and Welfare

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

$$V_{jsod} = \varepsilon_{jsd} w_{sd} Z_d^{-\gamma_s} h p_d^{-v_s} a_d \exp^{-M_{sod}} \quad (7)$$

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

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

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:

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

So the derivative of the CDF is given by:

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

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

$$\begin{aligned} \pi_{so1} &= \int \eta_s \varepsilon_{s1}^{-\eta_s-1} \exp\left\{-\left[\sum_{d=1}^D (\omega_{sod} \varepsilon_{sd})^{-\eta_s}\right]\right\} d\varepsilon_{s1} \\ &= \frac{1}{\sum_{d=1}^D \omega_{sod}^{-\eta_s}} \int \left(\sum_{d=1}^D \omega_{sod}^{-\eta_s}\right) \varepsilon_{s1}^{-\eta_s-1} \exp\left\{-\left[\varepsilon_{s1}^{-\eta_s-1} \left(\sum_{d=1}^D \omega_{sod}^{-\eta_s}\right)\right]\right\} d\varepsilon_{s1} \\ &= \frac{1}{\sum_{d=1}^D \omega_{sod}^{-\eta_s}} \int dF(\varepsilon) \\ &= \frac{1}{\sum_{d=1}^D \omega_{sod}^{-\eta_s}} \cdot 1 = \frac{(\widetilde{w}_{so1})^{\eta_s}}{\sum_{d=1}^D (\widetilde{w}_{sod})^{\eta_s}} \end{aligned} \quad (A.4)$$

The third line comes from the properties of the Frechet distribution, where we know that the term in the integral of the second line is simply the PDF with a shape parameter η , and a scale parameter $\sum_{d=1}^D \omega_{sod}^{-\eta_s}$. Expanding on the definitions for \widetilde{w}_{sod} , and scaling up the probability by the size of the skilled workforce P_{os} by origin, we derive labor supply by skill and destination:

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

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

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

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

Average city utility may depend on *hukou* costs. For instance, if a high-amenity city has a very restrictive *hukou* policy it may have a high average utility as those who originate from this city already have access to the amenities without paying *hukou* costs. We define average utility for those from city o to be:

$$\overline{V_{so}} \equiv \left(\sum_{d'} \left(w_{sd'} Z_{d'}^{-\gamma_s} h p_{d'}^{-v_s} a_{sd'} \exp^{-M_{sod'}}\right)^{\eta_s}\right)^{\frac{1}{\eta_s}} \quad (\text{A.6})$$

The equation shows that the average utility depends on the average option value migrating to any other city, and the ‘utility’ earned there. This average is scaled by the Frechet shape parameter η_s as it captures the dispersion in tastes across locations. The utility of those in city o is a decreasing function of 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 text, by using the above set of equations:

$$\log \pi_{sod} = -\eta_s \log \overline{V_{so}} + \eta_s (\log w_{sd} - v_s \log h p_d) + \eta_s \log a_{sd} - \eta_s \gamma_s \log Z_d - \eta_s M_{sod} \quad (\text{11})$$

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

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

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

Here the modified term $\left(\frac{1}{\rho}\right) (\log A_d + (1 - \rho) \log \bar{K}_d)$ 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 A.10, which includes an elasticity of substitution between labor ℓ_{sd} and skill biased capital k_{sd} represented by σ_y :

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

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

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

where ℓ_{sd} is the supply of workers of skill s , and L_{sd} is now a labor aggregate over workers and capital. Given this new set up, earnings can be represented by equation A.11, instead of equation 4:

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

D.3 A Note on Existence and Uniqueness of the Equilibrium

In Section 7, we present the model and equilibrium. Here we describe the determinants behind the existence and uniqueness of the equilibrium that we define in Section 7.3.

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.

E Additional Counterfactual Results

Table E1: The Productivity Effect of Reducing Pollution in Beijing (no externalities)

	Change in GDP per Worker (%)		
	Overall effects	Health	Relocation
	(1)	(2)	(3)
Reduce steady state PM2.5	10.364	5.819	4.295
Relax skilled hukou	4.931	0.000	4.931
Relax unskilled hukou	-4.654	0.000	-4.654
Reduce PM2.5 & relax skilled hukou	15.116	5.819	8.785
Reduce PM2.5 & relax unskilled hukou	5.707	5.819	-0.106

Notes: In this counterfactual exercise we reduce the steady state amount of pollution in Beijing by 50% (row 1). Next, we relax the *hukou* restrictions by skill level (rows 2 and 3) by 50%. Finally (rows 4 and 5) we lower the hukou regulation 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 through the pure relocation channel.

Table E2: The Productivity Effect of Relocating Pollution (no externalities)

	Change in GDP per Worker (%)		
	Overall changes	Health	Relocation
	(1)	(2)	(3)
Relocate steady state PM2.5	7.559	2.893	3.097
Relax hukou	2.318	0.000	2.318
Relax overall mobility constraints	6.319	0.000	6.319
Relocate PM2.5 & relax hukou	9.845	2.893	5.216
Relocate PM2.5 & lower migration costs	14.143	2.893	9.105

Notes: In this counterfactual exercise we relocate PM2.5 in all cities based on the amount of skill-biased capital in the city (row 1). In row 2, we relocate only the exogenous component of pollution. In addition to such relocations of pollution, we also relax the *hukou* restriction in the 24 top tier cities by 50% (row 2). In row 3 we relax overall migration costs to the 24 high tier cities by 50%. 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.

Table E3: 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: In this counterfactual exercise we reduce pollution according to the targets set by the 2013 *Air Pollution Prevention and Control Plan* (row 1). In addition to pollution regulations, we also relax the hukou restriction in higher tier cities (row 2), and migration costs (row 3) by 50%. Columns 1-3 show effects on skilled wages. Columns 4-6 on unskilled wages.