Childhood Exposure to Particulate Air Pollution, Human Capital Accumulation, and Income: Evidence from China

Avraham Ebenstein and Michael Greenstone*

Abstract

This paper examines the impact of childhood exposure to total suspended particulate (TSP) air pollution on human capital development and later labor market success. The analysis relies on a regression discontinuity design (RDD) generated by China's Huai River Policy, which provided subsidized coal for indoor heating during the winter to cities north of the Huai River but not to those to the south. The key findings are that children born just to the north of the Huai River were exposed to TSP concentrations that were about 50% higher, completed 0.83 fewer years of education, and earned 12%-22% lower incomes as adults, relative to children born just to the south. All findings are robust to using parametric and nonparametric estimation methods to adjust for distance from the Huai River, different kernel types, and alternative bandwidth sizes. The estimation of the income effects uses a RDD to quasi-experimentally vary TSP exposure, while also adjusting for all current city-level determinants of incomes by comparing migrants who are born just to the north and south of the Huai River.

* Greenstone: Department of Economics, University of Chicago, Chicago, IL 60637. Ebenstein: Department of Environmental Economics and Management, Hebrew University of Jerusalem, Rehovot 76100, Israel. Excellent research assistance was provided by Patrick Schwarz, Henry Zhang, Vishan Nigam, and Catherine Che. The authors declare that they have no relevant or material financial interests that relate to the research described in the paper.

I. Introduction

Particulate air pollution primarily results from the combustion of fossil fuels and has been a source of concern for human wellbeing since at least the 13th century in England when restrictions were put on burning coal. An extensive literature has documented a robust association between particulate air pollution and human health (Chen et al. 2013; Ebenstein et al. 2017; Pope et al. 2009), with infants found to be particularly at risk (Chay and Greenstone 2003; Currie and Neidell 2005; Currie et al. 2015). Indeed, some evidence suggests that particulate air

pollution is the greatest external risk to human health, even worse than cigarette smoking, dirty water, AIDs, etc. (Greenstone and Fan 2018).

Recently, there has been speculation and some evidence that particulates can affect human cognition, suggesting that the health effects literature may greatly understate the costs of particulate pollution. The more developed part of this literature suggests that fluctuations in particulate pollution affect worker productivity and student performance (Chang et al. 2016; Ebenstein et al. 2016; Schlenker and Walker 2016). Further, some important recent research suggests that early-life particulate pollution exposure results in permanently diminished cognitive ability (Isen et al. 2017; Bharadwaj et al. 2017). However, this literature has not had an opportunity to measure the causal effects of extended exposure to particulate air pollution on cognition, nor have there been opportunities to estimates its impacts at the very high concentrations that several billion people around the planet currently contend with.

In this paper, we exploit a novel quasi-experiment to examine the impact of childhood pollution exposure on adult human capital outcomes in China that addresses three challenges to currently understanding. First, China's Huai Winter Heating policy provides the basis for a regression discontinuity design (RDD) that produces plausibly exogenous permanent, rather than short-run, variation in childhood particulate air pollution exposure. Specifically, during the planning period (1950-1980), China initiated a policy that provided subsidized coal for indoor heating during the winter to cities north of the Huai River but not to those to the south (Almond et al. 2009; Chen et al. 2013; Ebenstein et al. 2017). This design provides a credible solution to the high potential for confounding air pollution with other determinants of human capital.

Second, China's household registration system places severe restrictions on migration. In fact, the paper focuses on the 1975 to 1982 birth cohort and for these people growing up outside one's designated address, or *hukou*, was nearly impossible (Zhao 2000). This is advantageous for our research design for two reasons. First, we can accurately assign an individual's childhood pollution exposure simply from knowing their *hukou* location. Second, under the *hukou* system, parents had little or no choice of where to reside, preventing selective migration from biasing our results. As such, the *hukou* system enables us to overcome two serious challenges that plague

many studies of pollution, where incomplete exposure data or selective migration makes it difficult to draw strong inferences.

Third, the Chinese context provides us with a large-scale sample of adults who were exposed to the elevated levels of pollution as children that prevail in many developing countries today. The existing literature on the relationship between cognition and pollution is both limited and carried out in developed countries with relatively modest levels of pollution (Isen et al. 2017; Bharadwaj et al. 2017). In the Chinese context, the policy dates back several decades and thus we are able to observed the people exposed as children at ages where people have typically completed schooling and are in the labor market.

Our empirical results demonstrate three key findings. First, the Huai River Policy generated a massive increase in pollution at the river, an effect that persisted for several decades. Using monitor-level data, across a set of specifications, we estimate that Northern residents were exposed to total suspended particulate (TSP) levels of 111-179 $\mu g/m^3$, or 39-63%, higher during the period 1981-2004. This implies the *increase* in pollution at the river is several times the level of allowed pollution in most developed countries currently (Kutlar et al. 2017), making this an ideal context to consider the consequences in adulthood of childhood exposure at the high levels that prevail in developing countries today. In more recent years, the monitoring system was enhanced to track PM_{10}, which are a subset of TSPs believed to be more detrimental to human well-being, and again we find that the increase in pollution at the Huai River is sizable; between 2003 and 2012, we estimate that PM_{10} levels were 27.0-35.8 $\mu g/m^3$, or 29-39%, higher in the North. Note however that for both periods, while we are able to demonstrate that pollution rises discontinuously at the Huai River, limitations in the monitoring system's geographical extent of coverage prevent us from being able to reliably assign pollution to China's entire population during the period that they were children. Therefore, we focus on the reduced form effect of growing up north of the Huai River, rather than the two-stage least squares estimate of the effect of TSPs, to understand the relationship between human capital accumulation and pollution exposure.

Our second finding is that children who grow up just north of the Huai River attain significantly lower levels of education than their counterparts born just south of the river. We estimate this by taking advantage of China's 2005 micro-census, which provides a nationally representative sample of individuals with detailed information on where they spent their childhood, where they live today, and their completed education. In our preferred specification, we find that children born north of the Huai River have 0.83 fewer years of education than children born just to the north. This finding is qualitatively robust to estimation by either parametric or non-parametric regression, varying our sample selection criteria, and is observed among both men and women. Further, in an examination of separate birth cohorts, we find that in the years of strictest enforcement of the Huai River Policy, the largest discontinuity in education is observed, lending further support to a causal interpretation of the connection between the policy and the decline in human capital at the river.

We then examine who in the educational distribution is affected by pollution exposure. We find that the decline in average education operates through a reduced proportion of children born just to the north of the Huai River graduating high school or pursuing post-secondary education as compared to children growing up just to the south. The implications of this finding are potentially enormous, as it implies that rampant air pollution in developing countries could be preventing their populations from reaching their full potential in terms of human capital. Indeed, it is difficult to identify intentional policies that raise or lower average education levels by 1 year.

Our third finding is that childhood pollution exposure significantly lowers adult incomes. Using parametric and non-parametric estimation methods, we estimate that growing up just north of the Huai River is associated with a 12-22% penalty in adult incomes. Besides leveraging the regression discontinuity design to adjust for determinants of human capital beside air pollution, the estimation of the income impact of childhood pollution requires us to tackle two additional empirical challenges. First, the higher level of pollution just to the north of the river may influence other determinants of incomes (e.g., the capital stock and local amenities), invalidating the RDD to isolate the effect of particulates holding all other factors (even those indirectly influenced by particulates) constant; that is, the RDD may be invalid as an approach to

recovering the partial derivative of incomes with respect to childhood particulate exposure. Therefore, we focus on specifications that include current city of residence fixed effects that adjust for all city-level determinants of incomes. The result is that the preferred specification is identified from variation in the incomes of workers born just to the north of the Huai River compared to workers born just to the south but working in the same city today. By the time of the 2005 Census, there was enough internal migration of adults in the labor force to produce sufficient variation so that this approach provides statistically meaningful estimates.

The second empirical challenge is that individuals do not migrate randomly and it is therefore important to consider the possibility of selective migration. For example, it is possible that the higher concentrations of particulate pollution reduce the returns to migrating by compromising individuals' human capital so that the migration rate is higher just to the south. A further possibility is that only the highest ability people from just to the north migrate due to lower returns to migration for people from the North. As a first effort to probe this possibility, we found that we were unable to reject the null of equal migration rates just to the north, relative to just to the south, using the RDD. Nevertheless, there could still be discrete differences in unobserved ability of migrants. For this reason, we also estimate the income models, including a Heckman-style correction to account for selection into migration. To identify this model, we exploit a strong 'first-stage' relationship between the distance to a major migrant destination (e.g. Beijing or Shanghai) from an individual's birthplace, and her tendency to migrate. Importantly, the adjustment for the inverse Mills ratio has little effect on the effect of birth just to the north of the river. This implies that even with our relatively demanding identification strategy, which is driven entirely by migrants who currently reside in the same location but differ only in where they were born, *and* for whom we are exploiting plausibly exogenous variation in childhood pollution exposure, we find that childhood pollution exposure significantly reduces adult earnings.

The rest of the paper proceeds as follows. Section II presents background information on the Huai River Policy and existing evidence on early childhood exposure to pollution. Section III presents our estimation strategy. Section IV describes our data and documents important trends. Section V presents our empirical results. We conclude in Section VI.

II. Background

A. China's Huai River Policy

During China's planning period (1950-1980), economic resources were allocated by government decision, and home heating was no exception. Government policy dictated that beginning in 1956 and implemented over the following decades, a coal-powered heating infrastructure would be established in all cities north of the boundary formed by the Huai River and Qinling mountain range. The residents of these cities were deemed to live in areas that merited subsidized or free home heating. This policy, however, had the unintended consequence of generating a large precipitous increase in air pollution at the Huai River by the burning of coal (Ebenstein et al. 2017). While the impact of the policy has declined over time as private sources of heating have become affordable and available, and other sources of pollution have become more relevant, the TSP increase at the Huai River in Chinese data from the 1980s is massive. The *increase* in TSP that we observe is several times the levels of TSP allowed under the United States Clean Air Act of 1970 and exceed even the acceptable levels according to China's own current guidelines (Chay et al. 2003; Huang et al. 2014). As such, this period of Chinese history represents a unique research opportunity to consider the health and human capital consequences of pollution exposure.

In this project, we focus on birth cohorts of the late 1970s and early 1980s who were exposed to the 'high tide' of the Huai River Policy. By this period, residents of northern cities enjoyed a fully subsidized and unified coal-powered home heating system, whereas residents of southern cities did not have the resources or the access to procure private sources of coal, so the difference in coal consumption was especially pronounced.[1] Another key aspect of the policy that we leverage is its focus on providing unified heating, which made the policy essentially an urban policy. Rural residents generally rely on individual (non-unified) heating methods, such as wood or coal powered stoves (*kang*), and so were largely unaffected by the policy.[2] As such, in our

[1] A firsthand description of this phenomenon may be found at: http://www.cma.gov.cn/kppd/kppdrt/201712/t20171215_458231.html

[2] Note that rural residents may have benefited from greater access to subsidized coal north of the Huai River, as the policy's impact would spill over into broader access to cheap coal for individual use. This could lead to pollution differences in rural areas north of the Huai River.

analysis we focus on the sub-population most directly affected by the Huai River Policy: children born in urban areas in the late 1970s and early 1980s (1975-1982).

B. Early-Life Exposure to Pollution and its Consequences

The potential health impact of fetal and early-life exposure to pollution has attracted a great deal of attention from both the medical field and from economists. Early scholarship focused on the impact of smoking during pregnancy on fetal health and birthweight (Fingerhut et al. 1990), but more recent work has also found that even ambient air pollution can affect infant health. Lowered birthweight, higher infant mortality, and long-term health problems have all been associated with fetal and early-life TSP exposure (Chay and Greenstone 2003; Currie and Neidell 2005; Currie and Walker 2011; Currie et al. 2015).

All PM_{10} particles, with diameters of no more than 10 micrometers (μm), are small enough to pass through the hairs in the nose and travel down the respiratory tract into the lungs, causing damage to lung cells. They can also pass into the lungs' alveoli, the blood vessel-covered air sacs in which the bloodstream exchanges oxygen for carbon dioxide. Through the alveoli, they enter the bloodstream. In addition, ultrafine particulate matter can be translocated directly to the brain via the olfactory nerve after inhalation (Oberdörster et al. 2004). However, since pollution operates through the circulatory system, it may affect cognitive development. Since the brain consumes a large fraction of the body's oxygen needs, any deterioration in oxygen quality can in theory affect cognitive development and performance (Clark and Sokoloff 1999; Calderón-Garcidueñas et al. 2008). Further, the first year of life because is particularly critical, since during this period the average child's brain doubles in size and the average child gains critical mental functions, including long term memory and speech (Bell and Fox 1994). Among multiple mechanisms by which particulate air pollution affects the brain, two that have emerged as important in the pathology and toxicology literature are those of inflammation and oxidative stress. In the various places in the body where particulate deposits, the lungs, the bloodstream, as well as the brain itself, the immune system's response to the intruding particles is inflammation. This means white blood cells release proteins called cytokines into the bloodstream. Cytokines are known to be linked to neurodegeneration, depression, and anxiety, such as through their

interaction with neurotransmitters (Miller et al. 2013). In addition to triggering inflammation, the particles contain soluble transition metal elements that cause oxidative stress, which is linked to aging, Alzheimer's, and other neurodegenerative diseases (Kelly 2003).

Prenatal and childhood exposure to air pollution is potentially most concerning, since the most critical stages of central nervous system development takes place in these years. For example, prenatal exposure to particulate matter has been found to be associated with increased carotid artery disease, a significant risk factor for cognitive impairment (Babadjouni et al. 2017). It may be a risk factor for the development of autism (Allen et al. 2017). Individuals exposed to higher particulate matter concentrations in childhood are at increased risk of developing neurodegenerative diseases such as Alzheimer's and Parkinson's (Calderón-Garcidueñas et al. 2013).

As a result of these physiological effects, a recent literature has been able to document that pollution significantly lowers labor productivity in a variety of contexts (Ham et al. 2011; Graff Zivin and Neidell 2012; Lavy et al. 2014; Chang et al. 2016). Scholars have also identified that long-term exposure to pollution is associated with decline in cognitive acuity among the elderly (Ailshire and Clarke 2015; Wilker et al. 2015). But the relationship between early-life pollution exposure and long-term human capital accumulation has not yet been examined rigorously in a developing country with extremely elevated pollution levels.[3]

III. Estimation Methods

A. Regression Discontinuity using the Huai River Policy

In our empirical strategy, we leverage the regression discontinuity (RD) design implicit in the Huai River Policy to measure its impact on long-term TSP and PM_{10} concentrations and human capital formation. This RD design exploits the Huai River Policy that provides free or heavily subsidized coal for indoor heating north of the river and no subsidies to the south. Specifically,

[3] Isen et al. (2017) find that early childhood exposure in the US during the early 1970s is associated with lower labor force participation and earnings at age 30. Bharadwaj et al. examine fetal exposure to air pollution on fourth grade test scores in Santiago, Chile, with a focus on carbon monoxide (Bharadwaj et al. 2017). Our paper focuses on TSP. In the Bharadwaj et al. study, mean PM_{10} levels in Santiago were 91$\mu g/m^3$, while PM_{10} levels in China have routinely exceeded 300 $\mu g/m^3$ in the past (Bharadwaj et al. 2017, Ebenstein et al. 2017).

we separately test whether the Huai River Policy caused discontinuous changes in TSPs by birth cohort and human capital to the north of the river. The respective necessary assumptions are that any unobserved determinants of TSPs or human capital change smoothly as they cross the river. If the relevant assumption is valid, adjustment for a sufficiently flexible polynomial in distance from the river or local linear regressions on either side of the river will remove all potential sources of bias and allow for causal inference.

In practice, we estimate the following parametric equations to test for the impacts of the Huai River Policy:

$$TSP_{ijt} = \alpha_0 + \alpha_1 N_j + f_0(L_j) + N_j f_1(L_j) + X_j\gamma + \lambda_t + u_{ijt} \qquad (1a)$$

$$Y_{ijt} = \delta_0 + \delta_1 N_j + g_0(L_j) + N_j g_1(L_j) + X_j\varphi + \tau_t + \varepsilon_{ijt} \qquad (1b)$$

for individuals i with *hukou* in county j born in birth cohort t. TSP_{ijt} is the total suspended particulate concentration assigned to each individual. N_j is an indicator variable equal to 1 for locations that are north of the Huai River line, $f(L_j)$ is a polynomial in degrees north of the Huai River that is interacted with N_j (chosen based on goodness of fit criteria), λ_t are year fixed effects and X_j is a vector of the demographic and county characteristics, other than air quality, that are associated with human capital. We then run the same regression with PM_{10} concentration in the place of TSP concentration.

An alternative estimation strategy for the RD approach involves non-parametric identification of (1a) and (1b). For example, consider the following set-up for estimation by local linear regression (Cleveland and Devlin, 1988):

$$TSP_{ijt} = \alpha_0 + \alpha_1 N_j + \alpha_2 L_j + \alpha_3 N_j L_j + \lambda_t + u_{ijt} \qquad (2a)$$

$$s.t. \ -h \le L_j \le h$$

$$Y_{ijt} = \delta_0 + \delta_1 N_j + \delta_2 L_j + \delta_3 N_j L_j + \tau_t + \varepsilon_{ijt} \qquad (2b)$$

$$s.t. \ -h \le L_j \le h$$

where h is latitude degrees of bandwidth (i.e., the acceptable distance from the discontinuity for sample inclusion). The choice of h involves balancing the conflicting goals of focusing comparisons near the river where the identification assumption is strongest and having a large

enough sample for reliable estimation. In the non-parametric results, we rely on a choice-rule which determines the optimal h as a function of the data. We also report results from the parametric RD approach described in equations (2a) and (2b) that restricts the sample to locations within 5 degrees latitude of the Huai River. This sample restriction which drops all observations that are distant from the Huai River is an informal way of implementing the local linear methods, which employ kernel weighting functions that assign greater weight to observations near the discontinuity.

B. Endogenous Modeling of Migration Decisions and Income Outcomes

In our empirical examination of the impact of the Huai River Policy on income, we focus our attention on migrants. Since variation in local income rates is driven by many factors unrelated to human capital, we include city of residence fixed effects to absorb this variation, and therefore our estimates are identified primarily from migrants. Since the decision to migrate is related to expected income, and pollution influences human capital, it is very likely that the choice of where to seek employment is itself influenced by pollution as well. Therefore, estimating the impact of pollution on income using migrants could be biased without accounting for this possible selection. We account for this selection in Heckman-style setup. Ideally, we are interested in estimating a model of the following form:

$$Y_{ijkt} = \delta_0 + \delta_1 N_j + \delta_2 L_j + \delta_3 N_j L_j + \tau_t + \eta_k + \mu_{1ijkt} \qquad (4a)$$

where Y_{ijkt} denotes log income of person i with birth *hukou* county j, residence county k, born in year t. τ_t is a birth-year fixed effect, and η_k is a residence county fixed effect. Note, however, we only observe Y_{ijkt} if person i chooses to migrate, which occurs when the benefits of migration outweigh the costs:

$$M_{ijkt} = 1 \leftrightarrow Y^*_{ijkt,dest} - Y^*_{ijkt,home} > f(Distz_{ij}) + \mu_{2ijkt} \qquad (4b)$$

where the worker can earn either Y^*_{dest} if she migrates or Y^*_{home} if she stays at her birthplace. Empirically, in our selection equation, we define $Distz_{ij}$ as a vector of distances from birth *hukou* county to six major employment destinations: Beijing, Tianjin, Shanghai, Guangzhou, Dongguan, and Shenzhen. This generates plausibly exogenous variation in the cost of potential migration (and therefore the probability of migrating). We treat other potential costs and benefits

of migration to an individual as unobserved and contained in μ_{2ijkt}. We follow the standard procedure of then including the selection term in our outcome equation.

$$Y_{ijkt} = \delta_0 + \delta_1 N_j + \delta_2 L_j N_j + \delta_2 L_j + \delta_3 N_j L_{j_j} + \tau_t + \eta_k + SEL + \mu_{3ijkt} \quad (4c)$$

The selection term SEL is calculated as the inverse Mills ratio from the first stage (4b).

IV. Data

Our main data set is China's 2005 Population and Housing survey, a nationally representative sample of roughly 0.2% of the population. Several features of this data set are desirable for this study. First, the sample is large, with 2.6 million observations overall, and roughly 88 thousand observations in our main study sample of all birth cohorts in 1975-1982 with urban *hukou*. Second, as a result of China's widespread migration since the late 1980s, care was taken in this survey to record each individual's county of birth and county of residence at a precise geographic level.[4] As such, we can precisely define the birthplace and early-life pollution exposure for all survey participants. This is particularly advantageous for our study since migration decisions are endogenous, and so it is critical to examine the human capital outcomes of all *births* occurring in a given county, rather than all current *residents*. Furthermore, in combination with the binding nature of the *hukou* system during China's central planning period, it is plausible that parents had little choice of where to raise their children, making a child's *hukou* exogenous to factors predicting parental quality. Third, China's population survey also measures education carefully in seven categorical variables, with education data available for all people in our sample born between 1975 and 1982. These features make China's 2005 population survey uniquely suited for analysing the relationship between early-life pollution exposure and adult human capital consequences.

In Table 1, we report our sample means of the overall census sample, and for our birth cohorts of interest. In Panel A, we present summary statistics for the 1975-1982 birth cohorts, who were exposed to the peak levels of enforcement of the Huai River Policy. The data for these cohorts in the 2005 survey are ideal for our study since they have (1) completed their education by this

[4] In most countries, birthplace is identified only at a higher level of aggregation. For example, US Census samples only include state of birth.

point and (2) were exposed to the peak levels of enforcement of the Huai River Policy. In particular, the urban born (columns 2 and 3) represent the subsample who would be most affected by the policy and so these 87,837 individuals will be our focus. We will focus primarily on the subset born within 5 degrees latitude of the Huai River, where our identification strategy is most compelling, which restricts the sample to 38,885 observations. We will perform robustness checks using the supplemented sample featured in Panel B, which reports statistics for all cohorts born between 1950 and 1974. These data will enable us to examine how the discontinuity in human capital observed at the Huai River evolved over time.

In Table 2, we report our sample means for our pollution readings for both TSP and PM_{10}. Our pollution data are taken from files published by China's Environmental Yearbooks, which are annual averages of daily mean readings of total suspended particulate matter less than 100 microns in width, or TSP, and particulate matter less than 10 microns in width, or PM_{10}. While our sample is to our knowledge the most complete account of air pollution readings in China for this period, we nevertheless face a significant challenge as a result of limited coverage during this period. For example, there were *no* TSP readings for the cohorts of the 1960s and 1970s, and only 36 stations were available in 1981. In the restricted sample of cities within 5 degrees of the Huai River, where our identification strategy is strongest, there are only 16 stations with readings. Although the monitoring system was expanded somewhat during the 1980s, assigning pollution readings to most of our population would require using stations quite distant from an individual's birthplace. As an alternative, we focus on demonstrating that pollution rises discontinuously at the Huai River, and we leverage our data's geographic precision on birthplace to draw inference from reduced form analysis.

During this period of extremely elevated pollution, China was undergoing radical changes in its educational distribution that would set the stage for its rise as an economic power. As shown in Figure 2, average education rose from roughly 9 years in the 1950 urban birth cohort to almost 13 years in the 1982 urban birth cohort. As a reference point, during this period average education in the United States stayed roughly constant at 13 years, meaning that urban China was able to essentially fully close the gap with the US in a relatively short period. The plot also highlights other noteworthy events during this period in Chinese history, including China's

Cultural Revolution of 1966, which brought China's education system to a virtual halt. Primary and middle schools gradually reopened after the first several months of the Cultural Revolution, but colleges and universities remained closed until 1970, with many premier institutions only reopened in 1972. The aftermath is readily apparent in the data as the cohorts born between 1962 and 1967 achieved successively lower levels of education than each previous cohort. As the impact of the Cultural Revolution subsided due to reforms enacted by Deng Xiaoping, education began to rise quickly. Children born in 1980 on average received 1.5 years of schooling more than their counterparts born one decade earlier and were around 20 percentage points more likely to acquire post-secondary education. The remarkable increase in education levels in China during this period and beyond continues to influence the global distribution of human capital. Today, China produces twice as many university graduates and eight times as many STEM graduates as the US.[5] And yet China arguably has yet to fulfill its full economic potential, especially in terms of research and development, an area in which it still lags behind the US.[6] But what might have been if not for China's severe pollution? Would China be in even stronger position globally in terms of innovation, without the harmful effects of air pollution? In the next section, we examine these questions using the natural experiment of the Huai River Policy, which we will demonstrate influenced pollution levels and human capital outcomes for these cohorts.

V. Empirical Results

A. Regression Discontinuity Estimates of the Huai River Policy's Impact on Pollution

In this section, we examine the impact of the Huai River Policy on pollution levels. A natural concern is that individuals would sort into either the north or south based on their preferences and the benefits of heating or costs of pollution (McCrary 2008). This issue is examined in Figure 3, where we report the population density in 0.2-degree latitude bins, overlaid with a plot generated by local linear regression to the south and north of the river. The results indicate that the major metropolitan area of Shanghai, which is 0.6 to 0.8 degrees to the south of the Huai

[5] https://www.forbes.com/sites/niallmccarthy/2017/02/02/the-countries-with-the-most-stem-graduates-infographic/#2ea3fd52268a

[6] https://phys.org/news/2016-03-world-leader-patent.html

River, is significantly larger than cities immediately to the north, influencing the plot and generating an apparent discontinuity in the population pattern. However, it is worth noting, that most bins on either side of the river are of relatively similar population. Challenges related to dealing with Shanghai are explored throughout the analysis, where we demonstrate the robustness of our core findings including or excluding Shanghai.

In Figure 4, we present a visual preview of our regression discontinuity estimates of the impact of the Huai River Policy on TSP and PM_{10} levels. The plot is generated using our pollution monitor data, restricted to stations with at least 10 years of data for each pollution measure. For each station, we calculate the long-term average of pollution, generating a sample of 636 station-years with TSP readings from 1981-2004 and 650 station-years with PM_{10} readings from 2003-2012. In the top panel, the plot reflects the dramatic increase in TSP at the Huai River during this period, a 172 $\mu g/m^3$ increase estimated by local linear regression. Interestingly, TSP appears to significantly increase with latitude north of the river but not south (coefficient of 52 $\mu g/m^3$ with standard error of 21 $\mu g/m^3$), suggesting that the subsidized coal was used in greater quantities in colder cities only in the north. PM_{10} levels also increase at the river; the estimated increase at the boundary is 28.5 $\mu g/m^3$. However, PM_{10} readings in the north decline with latitude, possibly owing to this form of pollution being less directly linked to heating and more related to other sources of pollution.[7] It is also likely that in the 2003-2012 period, other factors, such as emissions regulations, had greater influence on behavior and air quality.[8] Nevertheless, both plots present compelling evidence that cities just north of the Huai River were exposed to (and continue to be exposed to) significantly higher levels of particulate matter air pollution than those to the south.

In Table 3, we demonstrate that the finding of a discontinuous increase in TSP and PM_{10} at the Huai River is qualitatively robust to a rich set of alternative specifications. In our parametric specifications (columns 1-4), we estimate models in which our dependent variable is measured

[7] For example, PM_{10} emissions from heating sources is a mere 5% of the PM_{10} emissions from mining activities. This shows that PM_{10} pollution can be heavily influenced by industrial activities in a given area. http://www.npi.gov.au/npidata/action/load/emission-by-source-result/criteria/substance/70/destination/ALL/source-type/ALL/substance-name/Particulate%2BMatter%2B10.0%2Bum/subthreshold-data/Yes/year/2017

[8] It is also worth noting that between 1981-2004 and 2003-2012, private sources of home heating became more available, possibly mitigating the impact of the Huai River Policy, both in terms of the magnitude of the discontinuity and the relationship between cold weather and pollution.

pollution for a given station and year coefficient of interest is a North dummy variable, which represents the impact of the Huai River Policy on pollution. As control variables, we include year fixed effects, and a polynomial in the distance in latitude of the monitoring station from the Huai River. The polynomial is interacted with a North dummy, allowing the relationship between latitude and pollution to differ on either side of the river. We focus only on stations within 5 degrees latitude of the river in columns 1-3, thereby avoiding issues related to having estimates overly influenced by observations far from the discontinuity (Gelman and Imbens 2018). In column 4, we further restrict the sample to stations within 3 degrees latitude of the Huai River. We find that in all specifications except one, TSP rises at the river by a statistically significant amount, ranging from 124-179 $\mu g/m^3$. The exception is our third order polynomial, which includes 6 control variables for latitude terms and year fixed effects; it may be that since we have a large number of regressors relative to the number of observations, robust inference is more challenging.

As an alternative to parametric estimation, in columns 5 and 6, we present our estimates of the discontinuity using local linear regression. Our preferred estimate is reported in column 5, where we use local linear regression with an Epanechnikov weighting kernel and a 5-degree bandwidth, and estimate a 172 $\mu g/m^3$ increase in TSP. This estimate declines to 111 $\mu g/m^3$ when the bandwidth is chosen using an algorithm that minimizes the mean squared error (MSERD), chosen to be 2.88 degrees in this particular sample (Calonico et al. 2014). The results for PM_{10} are similarly conclusive that the Huai River generates a discontinuity in pollution. Using the same methodology as with TSP, we find that the parametric estimates for PM_{10} are generally statistically significant but are sensitive to specification choice. We instead favor the estimates produced by local linear regression find that PM_{10} increases by 28.5 and 27.0 $\mu g/m^3$ in the 5-degree and MSERD sample respectively, both statistically significant and precisely estimated (columns 5 and 6).

B. Regression Discontinuity Estimates of the Huai River Policy's Impact on Education

In light of the steep increase in pollution at the Huai River, we turn to a parallel examination of the discontinuity in educational attainment at the boundary. In Figure 5, we present a plot of

average educational attainment in 0.2-degree latitude bins for the region to the south and north of the Huai River. Each bubble is sized proportionally to the population living in that region, and the figure is overlaid with local linear regressions to each side of the Huai River. As shown in the plot, a significant discontinuity of around 0.9 years of education is observed for cohorts born in our sample period. The results are most compelling very near the river, such as cities within one degree of latitude, where the cities just south of the river appear to have significantly higher educational attainment than those just to the north. As a robustness check, we also run the analysis without Shanghai, which is located just south of the river, and find a statistically significant discontinuity of 0.5 years.

These results are complemented by those in Table 4, where we examine the discontinuity in education at the Huai River using a set of specifications parallel to those used in Table 3 to examine the discontinuity in pollution. Note that our sample of births and their educational outcomes is much richer than our pollution measurement data, and we have 38,885 observations meeting our sample selection criteria. This leads to very precise measures of the decline in education at the Huai River in both our parametric and non-parametric specifications, ranging from a 0.66-1.17 estimated decline in years of education at the boundary. Relative to southern average education in these cohorts of 12.4 years, this implies that they suffered between a 5% and 9% diminution in human capital. Our preferred estimate is reported in column 5; using local linear regression in our 5-degree sample, we estimate that the decline in education at the Huai River is a precise 0.83 years, a striking finding suggesting that children born just north of the river were significantly less academically successful than those born to the south. Interestingly, when the results are broken out separately for men and women, there is an incredible consistency in the findings. While our estimate of the decline at the river is somewhat larger for men than women (-0.87 versus -0.79), the general consistency of the results across gender in Table 4 is consistent with an interpretation that greater pollution exposure north of the river is similarly harming female and male educational outcomes.

In Figure 6, we perform this same calculation but for all birth cohorts from the preceding four decades. Using our expanded sample of cohorts born 1950-1982, we estimate the discontinuity in education by year with our preferred specification (local linear regression in the 5-degree

sample). Interestingly, we find that the birth cohorts for which we estimate the largest decline in education at the river roughly matches our understanding of when the policy was most relevant for determining air pollution. In particular, the decline in education at the Huai River is evident but modest in the late 1950s when the policy newly announced, is negligible during the 1960s (perhaps because these birth cohorts affected by the Cultural Revolution during their schooling years were limited to few years of schooling regardless of their aptitude), and then becomes larger for each successive birth cohort in the late 1970s when the policy was fully implemented and subsidized heating provided. This suggests that the combination of increasing educational opportunities for the brightest students, and the worsening damage from air pollution, combined to make the negative consequences of the Huai River Policy grow increasingly large throughout the 1970s, achieving a maximal level of influence on education for our cohorts of interest (1975-1982).

We examine mechanisms for these findings in Figure 7. As mentioned earlier, our variable on years of education is constructed from categorical responses in the census questionnaire. In Figure 7, we plot the change in density of each category at the Huai River. The results indicate that children born north of the Huai River have a significantly higher chance of only receiving 7-9 years of schooling but are far less likely to pursue post-secondary education. In particular, we find an 7.6% and 1.8% decline in the proportion reporting 13-16 years of education and post-BA education (16 years or more) respectively. Relative to a baseline proportion of 41.4% reporting matriculation in post-secondary education south of the Huai River, this combined decrease represents a 23% decrease in the total fraction of students attending higher education north of the Huai River.

In Figure 8, we explore a natural placebo exercise by examining whether the discontinuity observed at the Huai River is found at displacements of the Huai River. We perform the exact same local linear regression analysis as in our preferred specification from Table 4 but using one-degree displacements of the Huai River. This is executed 11 times, using the true Huai River, 5 displacements north, and 5 displacements south of the river. Notably, the Huai River discontinuity is the largest in magnitude. The plot also indicates that there are some positive and negative discontinuities elsewhere. These are harder to explain. Although this placebo's test

results are not the most decisive that we have encountered in our careers, these results broadly support the validity of the Huai River research design.

We further examine the decline in education at the Huai River by examining the impact of the policy by prefecture in Figure 9. This is executed with a local linear regression where years of education are regressed on distance to the river, the interaction of distance with a north indicator, and birth year fixed effects, and all these coefficients are constrained to have the same values as in the regression for Column 5 of Panel A of Table 4, then plotting the prefectural means of the residuals (incorporating the discontinuity at the river). The figure reveals that the phenomenon of reduced education for children born north of the river is not being driven by a few outlier prefectures, but rather most prefectures contribute to the negative discontinuity. The figure demonstrates that the impact of the Huai River Policy on education is felt broadly across many cities in the region just north of the river.

C. Regression Discontinuity Estimates of the Huai River Policy's Impact on Adult Incomes

In this section, we examine the impact of the Huai River policy on adult earnings. This is a natural exercise to perform in light of the large decline in education observed at the river and the well-established relationship between education and earnings (Card 1999). The effect of particulates on incomes could operate through both lower levels of education and through impaired cognition, even conditional on years of education, so we think this is unlikely to be a valid setting to estimate the returns to education.

We begin our analysis by estimating the discontinuity in adult incomes at the Huai River in models without including fixed effects for the worker's current city of residence (and presumably employment). These specifications are unlikely to isolate the effect of particulate pollution because the pollution differences north and south of the river are likely to cause other differences in the determinants of income. For example, the lower education levels seem likely to alter the capital-labor ratio, amenity levels, and land prices (Roback 1982; Chay and Greenstone 2003; Currie et al. 2015. In this respect, these specifications are well suited to infer the total effect of the Huai River Policy (i.e., the total derivative of a change in pollution) but poorly

suited to isolate the effects of childhood exposure to air pollution (i.e., the partial derivative of a change in pollution). Indeed, in these models without resident city fixed effects, we find quite large income effects in the 5-degree sample: naively regressing growing up north of the Huai River on income is associated with an income penalty of roughly 50 logarithmic points using both parametric and non-parametric estimation methods. In column 3, in a more limited sample chosen by MSE-optimal bandwidth RD (Calonico et al. 2014), the discontinuity in income at the Huai River is smaller, about 25 percent.

Our aim is to understand the effect of childhood particulates exposure on income, while holding other factors constant. Thus, columns 4-6 report on specifications that include current residence city fixed effects that adjust for all city-level determinants of income, including current pollution concentrations. Here, the effect of the Huai River Policy is identified from differences in income among migrants born just to the north and south of the river. This model can be estimated with migrants only, although we report results from fitting the model on the full sample of both migrants and non-migrants, as the non-migrants help to identify the model's other parameters.

This strategy absorbs considerable variation, and our estimates of the income impacts are now in a tighter band between 18 and 22 percent and are much more precisely estimated. The estimated income penalty is larger than the conventional wisdom about the returns to an additional 0.8 years of education in the US, but, as we noted, it is possible that particulates impair cognition, conditional on education. Additionally, there is a much less well-developed literature on the returns to education in China that these estimates can be benchmarked against. Notably, there is little difference in the effects for men and women, suggesting that pollution is similarly harmful to both men and women.

The reliance on migrants for identification of the income impact of growing up north of the Huai River generates a new empirical challenge, however. Who migrates and why? This is potentially related to human capital, and pollution exposure itself. For example, childhood exposure to particulate pollution might reduce the returns to migrating by compromising individuals' human capital so that the migration rate is higher just to the south. As such,

modeling the decision to migrate is critical to estimation. As a first pass to understand the degree to which selection into migration poses an empirical challenge, we first examine whether the probability of migrating changes at the Huai River.

Table 6, reports on the estimation of parametric and non-parametric RDD models where the dependent variable is a dummy variable whether the worker left her prefecture (Panel A) or province (Panel B). Reassuringly, there is little evidence of discrete changes in the probability of migration among individuals born just to the north and south of the river.

A subtler possibility is that there are differences in unobserved skills of migrants among those born just to the north, relative to those born just to the south. For example, although the total migration rate is equal, it might be that more high and low skilled people migrate from the North and more middle-skilled people migrate from the South. We therefore estimate a 2-step Heckman-style selection correction approach that aims to adjust the estimates of the Huai River Policy's impacts for differences in unobserved determinants of income by explicitly modeling the migration decision. As we discussed above, this model is identified by excluding individuals' distances from key migration destinations from the wage equation.

Table 7 present estimates of the impact of the Huai River Policy on income that are corrected for potential endogeneity of the migration decision. Column 1 is our baseline estimate and the specification is similar to the one in Column 5 of Table 5; the only difference is that people whose residence county only contains one observation in our 0.2% sample are dropped from all Table 7 regressions for computational reasons, due to the residence county fixed effects. This specification suggests that workers suffer a roughly 20 percent income penalty growing up north of the river, in line with the Columns 4-6 estimates in Table 5. In column 2, the same specification is fit on a sample of migrants *only*, defined to be people whose residence prefecture differs from their *hukou* prefecture. Migrants comprise roughly a fifth of the sample (4,756 workers). The estimate is nearly identical to that produced in column 3, where we find a 19% wage penalty from growing up north of the river.

We turn in columns 3-9 to our selection-corrected estimates, which are estimated in a two-step Heckman-style procedure. The first stage involves predicting migration using a vector of distances from one's birthplace to six primary migrant destinations: Beijing, Dongguan, Guangzhou, Shanghai, Shenzhen, and Tianjin. In columns 3-8, we take turns including the distance to each of these cities separately and report the probit coefficients. None of these specifications is our focus, but it is apparent that the probability of migration to any city is declining in distance from each of these cities.

Our focus is the much richer equation that includes the distances to each of these cities in the migration equation for a total and these results are reported in Column 9. It is more challenging to interpret any one of the distance coefficients because now they are all conditioned on each other, but it is again evident that distance affects the probability of migration. The key finding from this specification, and indeed from the table, is that after adjustment for the selection correction term, the effect of being born just north of the rivers, relative to just south, is to reduce income by about 12%. This is smaller than the unadjusted estimate in column 2, but the 95% confidence intervals of the two estimates certainly overlap. Thus, the effect of childhood exposure to particulate pollution remains negative and economically meaningful after adjustment for the Heckman selection correction term derived from a migration equation. Further, it is noteworthy that the transformation of the rho coefficient (i.e., *arthrho*) reveals that there is positive selection in migration; that is, the individuals who are more likely to migrate after conditioning on covariates are also likely to have higher income than predicted by their covariates.

VI. Conclusions

This paper has examined the relationship between early childhood pollution exposure and human capital outcomes in China. We demonstrate three key results. First, the Huai River Policy, which dictated that northern Chinese cities would be provided with unified coal-driven heating, helped contribute to the dangerously high levels of particulate air pollution in northern China. While the policy had good intentions regarding keeping residents warm in the winter, its unintended consequence was to expose residents of cities just north of the river to severely worse

pollution than those just south of the river. Previous work has demonstrated the substantial negative impacts on health, particularly life expectancy (Ebenstein et al. 2017).

Our second finding is that the Huai River Policy significant decreased educational attainment among northern residents relative to their southern counterparts. We estimate that the policy reduced educational attainment by nearly a full year, and these effects are manifest in a lower fraction of northern children participating in post-secondary education. The decline in education at the Huai River is striking and suggests that childhood exposure to particulate pollution can have powerful and harmful effects on the long-term cognitive ability of children.

Our third finding is that the Huai River Policy also resulted in a decline in adult income for the exposed birth cohorts. We estimate that children who grew up just north of the Huai River suffered a 12-22 percent decline in adult income, relative to those who grew up just to the south. This result is identified from variation in income among workers living in the same city today who migrated from different parts of China. Further, the basic finding is robust to estimating a two-stage Heckman model where we endogenously model the decision to migrate and we correct for selection in the wage equation. This novel approach produces compelling evidence that the long-term consequences of pollution are manifest in adult income.

Today, there are hundreds of millions of children in developing countries growing up in cities with pollution levels similar to those found in China during our period of analysis. In the context of the debate regarding the benefits and costs of stricter environmental regulation, a narrow focus on the human health cost of particulate matter air pollution likely understates the true cost to society of particulate pollution. However, the magnitude of the results and their import for social welfare mean that more research would be of high value.

VII. References

Ailshire, Jennifer, and Philippa Clarke. 2015. "Fine Particulate Matter Air Pollution and Cognitive Function Among U.S. Older Adults." *Journal of Gerontology B Psychological Sciences 70*(2): 322-328.

Allen, J. L., G. Oberdorster, K. Morris-Schaffer, C. Wong, C. Klocke, M. Sobolewski, ... & D. A. Cory-Slechta. 2017. "Developmental neurotoxicity of inhaled ambient ultrafine particle air pollution: parallels with neuropathological and behavioral features of autism and other neurodevelopmental disorders." *Neurotoxicology 59*: 140-154.

Almond, Douglas, Yuyu Chen, Michael Greenstone, and Hongbin Li. 2009. "Winter Heating or Clean Air? Unintended Impacts of China's Huai River Policy." *American Economic Review Papers & Proceedings 99*(2): 184-90.

Babadjouni, R. M., D. M. Hodis, R. Radwanski, R. Durazo, A. Patel, Q. Liu, & W. J. Mack. 2017. "Clinical effects of air pollution on the central nervous system: a review." *Journal of Clinical Neuroscience 43*: 16-24.

Bell, M.A. and N. A. Fox. 1994. "Brain development over the first year of life." *Human Behavior and the Developing Brain*: 314-345.

Bharadwaj, P., M. Gibson, J.G. Zivin, and C. Neilson. 2017. "Gray matters: fetal pollution exposure and human capital formation." *Journal of the Association of Environmental and Resource Economists 4*(2): 505-542.

Calderón-Garcidueñas, L., M. Franco-Lira, A. Mora-Tiscareño, H. Medina-Cortina, R. Torres-Jardón, & M. Kavanaugh. 2013. "Early Alzheimer's and Parkinson's disease pathology in urban children: Friend versus Foe responses--it is time to face the evidence." *BioMed Research International 2013*: 1-16.

Calderón-Garcidueñas, L., A. Mora-Tiscareño, E. Ontiveros, G. Gómez-Garza, G. Barragán-Mejía, J. Broadway, S. Chapman, G. Valencia-Salazar, V. Jewells, R. Maronpot, C. Henríquez-Roldán, B. Pérez-Guillé, R. Torres-Jardón, L. Herrit, D. Brooks, N. Osnaya-Brizuela, M. Monroy, A. González-Maciel, R. Reynoso-Robles, R. Villarreal-Calderon, A. Solt, and R. Engle. 2008. "Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs." *Brain and Cognition 68*(2): 117-127.

Calonico, S., M. D. Cattaneo, and R Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica 82*: 2295–2326.

Card, David. 1999. "Chapter 30 - The Causal Effect of Education on Earnings." *Handbook of Labor Economics 3*: 1801-1863.

Chang, Tom, Joshua Graff Zivin, Tal Gross, Matthew Neidell. 2016. "Particulate Pollution and the Productivity of Pear Packers". *American Economic Journal: Economic Policy 8*(3): 141-69.

Chay, Kenneth, and Michael Greenstone. 2003. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics 118*(3): 1121-1167.

Chay, Kenneth, Carlos Dobkin, and Michael Greenstone. 2003. "The Clean Air Act of 1970 and Adult Mortality." *Journal of Risk and Uncertainty 27*(3): 279-300.

Chen, Yuyu, Avraham Ebenstein, Michael Greenstone, and Hongbin 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 110*(32): 12936-12941.

Chen, Z., J.N. Wang, G.X. Ma, and Y.S. Zhang. 2013. "China tackles the health effects of air pollution." *The Lancet* 382(9909): 1959-1960.

Clark, D. and L. Sokoloff. 1999. "Circulation and energy metabolism of the brain," in G. Siegel, B. Agranoff, R. Albers, S. Fisher and M. Uhler, eds., *Basic Neurochemistry*: *Molecular, Cellular and Medical Aspects*. Lippincott-Raven: 637-670.

Cleveland, William, and Susan Devlin. 1988. "Locally weighted regression: an approach to regression analysis by local fitting." *Journal of the American Statistical Association 83*(403): 596-610.

Currie, Janet, and Matthew Neidell. 2005. "Air pollution and infant health: What can we learn from California's recent experience?" *Quarterly Journal of Economics 120*(3): 1003-1030.

Currie, Janet, and Reed Walker. 2011. "Traffic congestion and infant health: Evidence from EZPass." *American Economic Journal: Applied Economics 3*(1): 65-90.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." *American Economic Review 105*(2): 678-709.

Dominici, F., M. Greenstone, and C. Sunstein. 2014. "Particulate Matter Matters." *Science 344*(6181): 257–259.

Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng 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.

Gelman, Andrew and Guido Imbens. 2018. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*: 1-10.

Greenstone, M. and C.Q. Fan. 2018. "Introducing the Air Quality Life Index." Retrieved from https://aqli.epic.uchicago.edu/reports/.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review 102*(7): 3652-3673.

Fingerhut, L., J. Kleinman, and J. Kendrick. 1990. Smoking before, during, and after pregnancy. *American Journal of Public Health 80*(5): 541-544.

Ham, John, Jacqueline Zweig, and Edward Avol. 2011. "Pollution, Test Scores and the Distribution of Academic Achievement: Evidence from California Schools 2002-2008." *IZA Working Paper*.

Huang, R.J., Y. Zhang, C. Bozzetti, K.F. Ho, J.J. Cao, Y. Han, K.R. Daellenbach, J.G. Slowik, S.M. Platt, F. Canonaco, and P. Zotter. 2014. "High secondary aerosol contribution to particulate pollution during haze events in China." *Nature 514*(7521): 218-222.

Isen, Adam. 2017. "Every Breath You Take—Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970." *Journal of Political Economy* 125(3): 848-902.

Kelly, F.J. 2003. "Oxidative stress: Its role in air pollution and adverse health effects." *Occupational and Environmental Medicine, 60:* 612-616.

Kutlar, J. M, M. Eeftens, E. Gintowt, R. Kappeler, N. Kunzli. 2017. "Time to Harmonize National Ambient Air Quality Standards." *International Journal of Public Health* 62(4): 453–462.

Lavy, Victor, Avraham Ebenstein, and Sefi Roth. 2014. "The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation." *NBER Working Paper* 20648.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics 142*(2): 698-714.

Miller, A. H., E. Haroon, C. L. Raison, and J. C. Felger, 2013. "Cytokine targets in the brain: impact on neurotransmitters and neurocircuits." *Depression and Anxiety* 30(4): 297-306.

Oberdörster, G., Z. Sharp, V. Atudorei, A. Elder, R. Gelein, W. Kreyling, and C. Cox. 2004. "Translocation of inhaled ultrafine particles to the brain." *Inhalation Toxicology 16*(6-7): 437-445.

Pope, C. A., M. Ezzati, and D. Dockery. 2009. "Fine-Particulate Air Pollution and Life Expectancy in the United States." *New England Journal of Medicine 360*: 376-386.

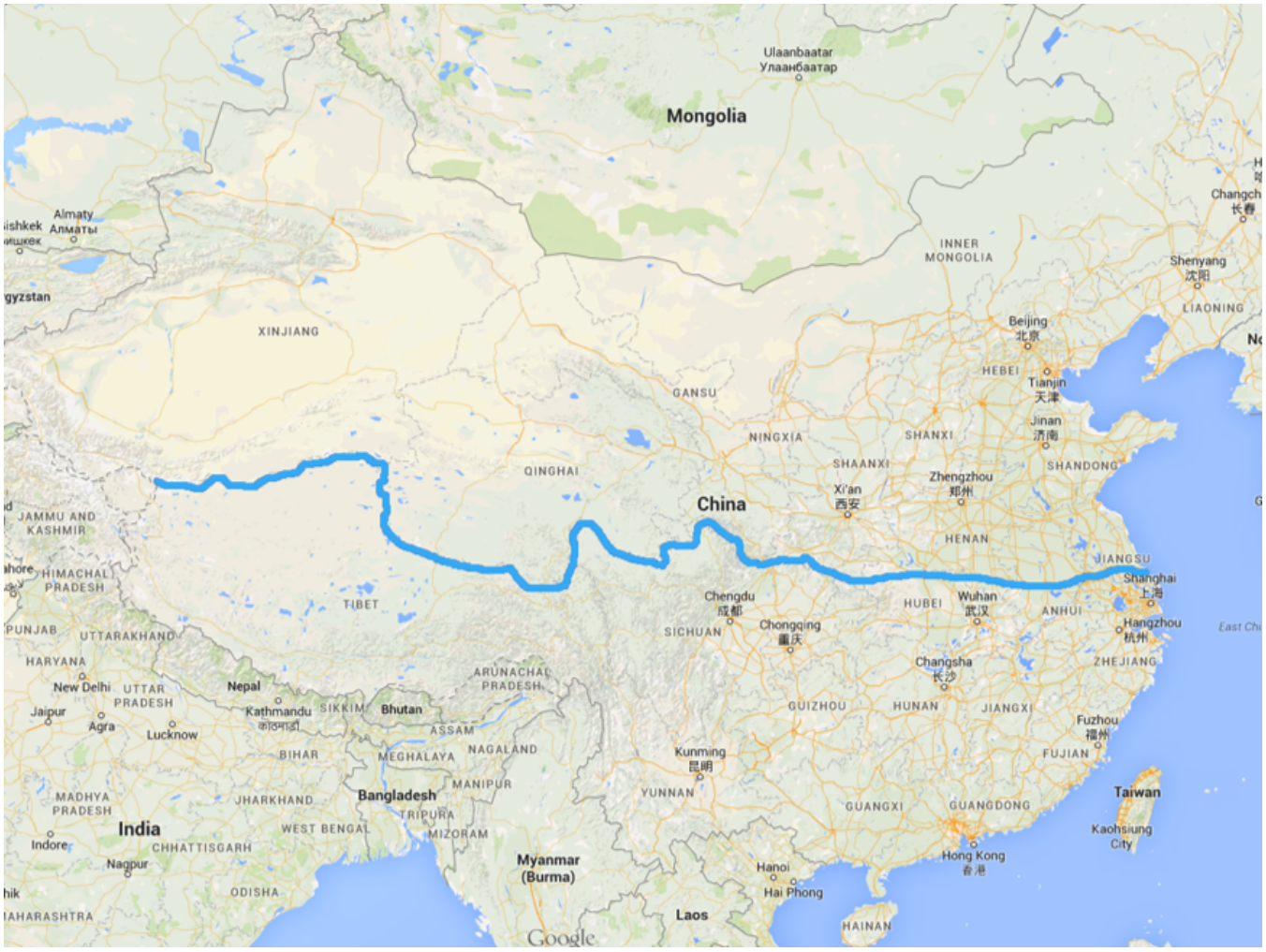
Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6):1257-1278.

Schlenker, Wolfram, and Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *Review of Economic Studies 83*(2): 768-809.

Wilker, E. H., Preis, S. R., Beiser, et al. 2015. "Long-Term Exposure to Fine Particulate Matter, Residential Proximity to Major Roads and Measures of Brain Structure." *Stroke 46*(5): 1161-1166.

Zhao, Yaohui. 2000. "Rural-to-urban labor migration in China: the past and the present," in Loraine A. West and Yaohui Zhao, eds., *Rural Labor Flows in China*. Berkeley Institute of East Asian Studies, University of California: 15-33.

Figure 1: The Huai River Dividing Line



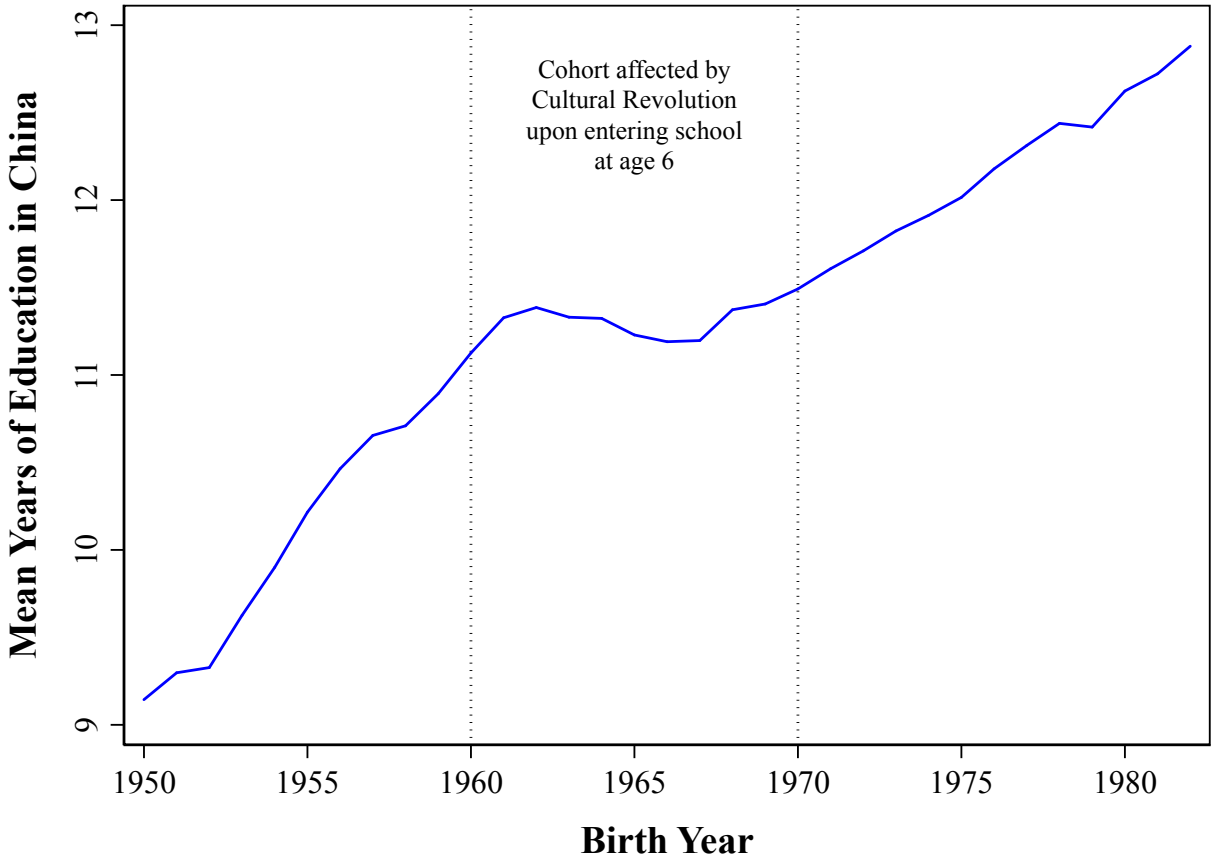
Notes: The Huai River, along with the Qinling Mountains and the Tibet border to the west of the headwaters of the Huai River, forms the dividing line between regions with heavy coal-fired home heating and regions with little to no use of such boilers, as the result of the planning-era Huai River Winter Heating Policy.

Table 1: Summary Statistics By Birth Cohort

		Urban		Rural	
	All	Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)
Panel A: 1975-1982 Birth Cohorts					
Minority	10.3%	6.4%	6.5%	12.3%	11.6%
Highest Level of Education Attempted					
No Schooling	2.3%	0.3%	0.3%	2.0%	4.0%
Primary School	14.9%	1.9%	2.4%	16.7%	23.3%
Middle School	53.6%	24.3%	25.8%	67.5%	63.1%
High School	16.7%	33.7%	33.0%	11.7%	8.2%
Vocational College	8.1%	23.8%	24.7%	1.7%	1.3%
Baccalaureate College	4.1%	14.7%	12.5%	0.3%	0.2%
Post-Graduate	0.4%	1.3%	1.3%	0.0%	0.0%
Years of Education	9.66	12.50	12.36	8.79	8.28
	(3.08)	(2.83)	(2.86)	(2.22)	(2.47)
Monthly Income (RMB)	580	1,006	729	614	331
	(723)	(1072)	(880)	(635)	(398)
Observations	288,111	41,640	46,197	96,245	104,029
Panel B: 1950-1974 Birth Cohorts					
Minority	7.8%	5.6%	5.4%	8.9%	8.6%
Highest Level of Education Attempted					
No Schooling	6.7%	0.6%	1.9%	4.1%	13.3%
Primary School	29.6%	6.3%	10.1%	32.1%	43.8%
Middle School	43.6%	35.9%	39.1%	53.7%	38.7%
High School	13.8%	31.8%	31.0%	9.6%	4.0%
Vocational College	4.3%	16.2%	12.4%	0.4%	0.1%
Baccalaureate College	1.9%	8.3%	5.1%	0.0%	0.0%
Post-Graduate	0.2%	0.9%	0.4%	0.0%	0.0%
Years of Education	8.33	11.34	10.59	7.98	6.62
	(3.39)	(3.01)	(3.16)	(2.50)	(3.07)
Monthly Income (RMB)	521	1,004	561	552	282
	(736)	(1142)	(844)	(678)	(313)
Observations	1,035,373	163,352	156,549	352,905	362,567

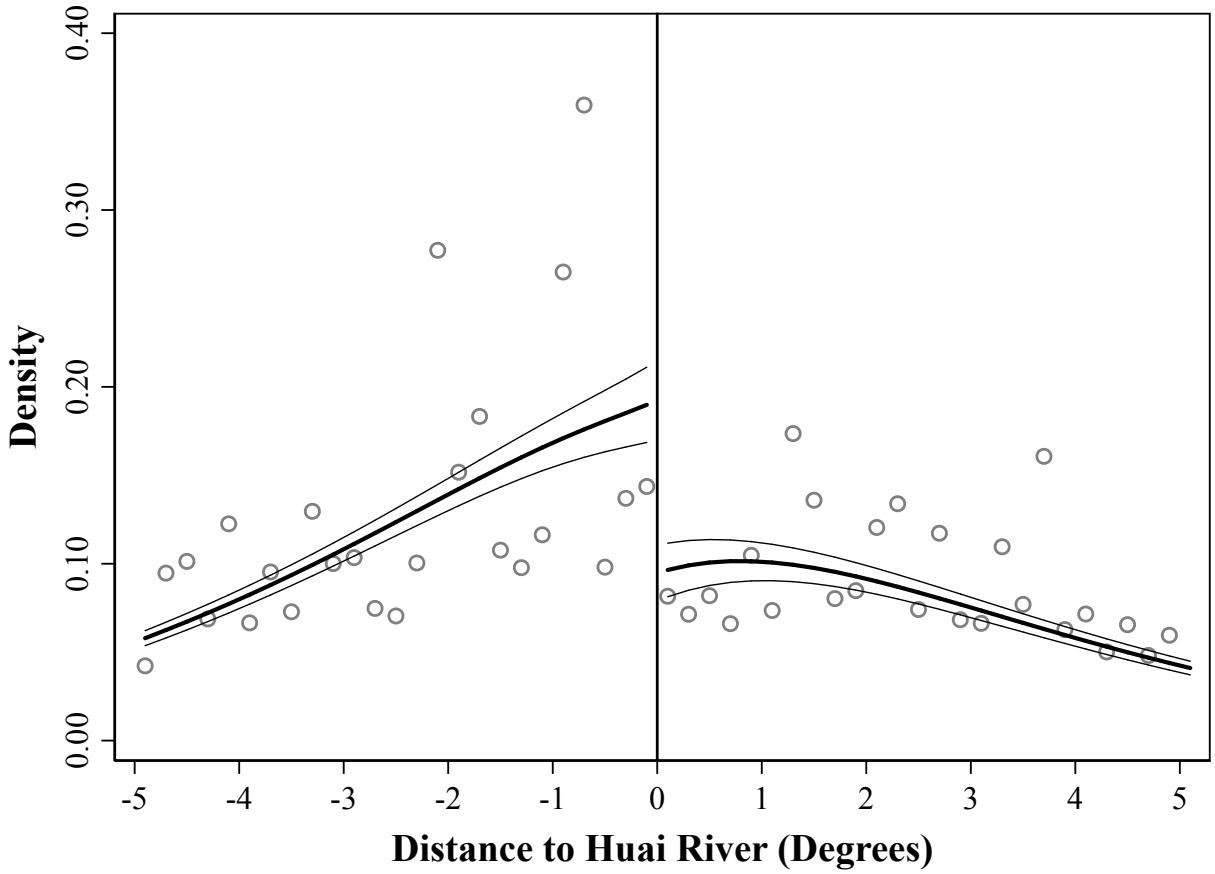
Notes: Data are from a random 20% sample of the 1% microdata of the 2005 Chinese census. Statistics are computed for all Chinese people in the given birth cohort range and with the given characteristics.

Figure 2: China's Remarkable Increase in Educational Attainment



Notes: This graph displays the mean years of education in urban China by birth cohort, weighted by sampling weights. The dotted vertical lines show the beginning and end of the birth cohorts exposed to the Cultural Revolution (1966-1976) at age 6, upon entering primary school.

Figure 3: McCrary Selection Test



Notes: This figure displays the McCrary (2008) selection figure, using the five-degree sample for urban *Hukou* holders in the 1975-1982 birth cohorts. The McCrary selection test was modified to implement local linear regression with Epanechnikov kernel and sampling weights. Standard errors are clustered on *Hukou* county. Each bubble's y-value displays the relative population of a 0.2-degree bin for birth *Hukou* distance to the Huai River. To obtain each bin's proportion of the total sample in the five-degree sample, multiply the density value by 0.2.

Table 2: Pollution Monitors by Year

	Number of Monitors		Mean Reading ($\mu g/m^3$)	
	North	South	North	South
Panel A: TSP Air Polluton Monitors				
Full Sample				
1981	23	13	926	395
1990	65	66	498	278
2000	48	46	336	186
<5 Degrees of Huai				
1981	6	10	835	380
1990	21	35	504	305
2000	19	27	356	210
Panel B: PM10 Monitors				
Full Sample				
2003	52	54	140	102
2007	57	56	105	83
2012	162	163	86	66
<5 Degrees of Huai				
2003	30	35	141	113
2007	30	36	105	93
2012	75	87	91	75

Notes: Annual averages of daily mean readings of air pollution data are from China?s national pollution monitoring system, via the Environmental Yearbooks, supplemented with data from the World Bank and the Chinese Environmental Protection Agency. Total suspended particulates (TSP) consist of particulate matter under 100 microns in width, while PM_{10} consist of particulate matter under 10 microns in width.

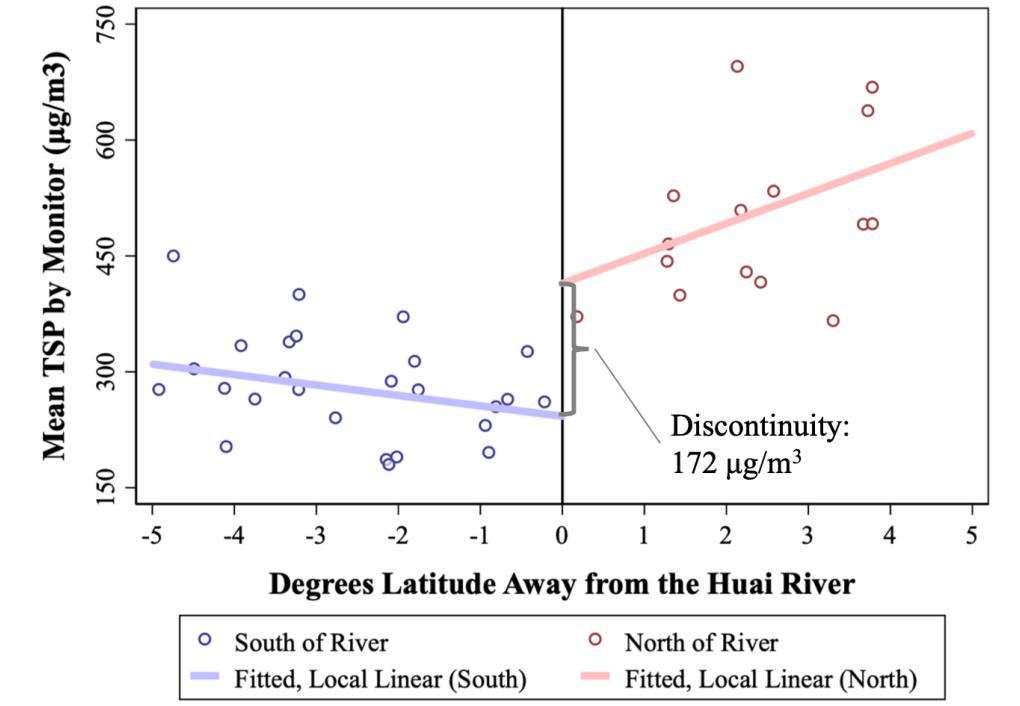
Table 3: Pollution Discontinuity Table

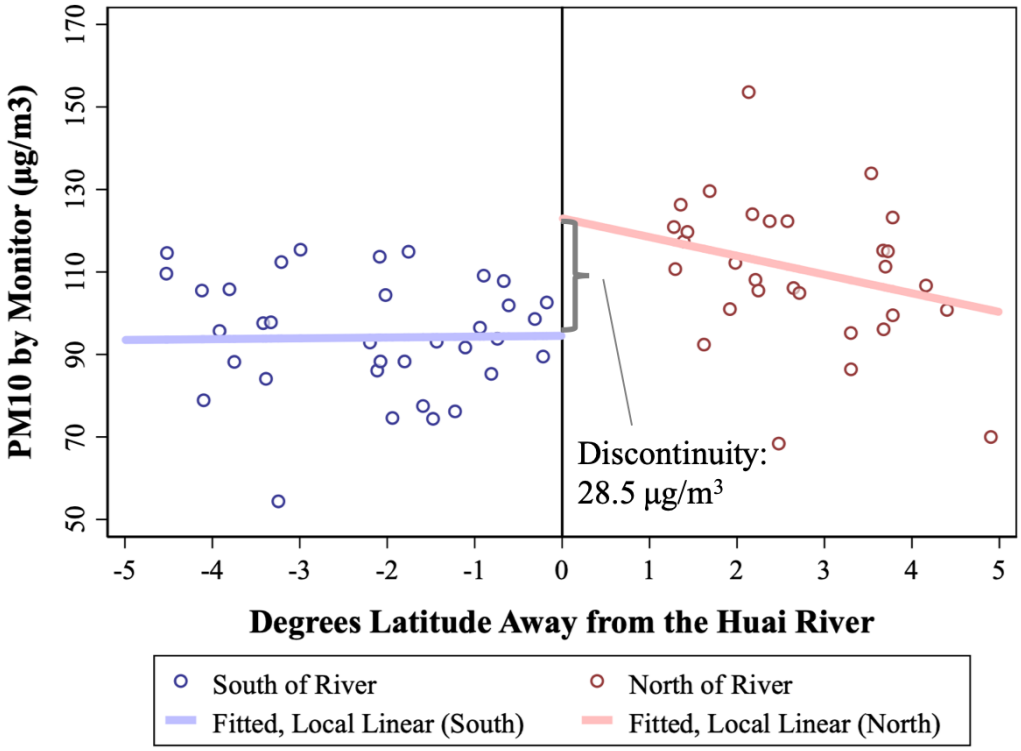
	Polynomial				Local Linear	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: TSP Air Pollution Monitors, 1981-2004						
TSP (μg/m3)	179***	135**	28	124**	172***	111**
	(52)	(64)	(67)	(50)	(46)	(47)
Num Obs.	636	636	636	389	636	389
Bandwidth	5-Deg	5-Deg	5-Deg	3-Deg	5-Deg	3-Deg
Mean TSP South of Huai						
1981-2004	282	282	282	256	271	259
Panel B: PM10 Monitors, 2003-2012						
PM10 (μg/m3)	33.9***	7.6	88.2*	35.8***	28.5***	27.0**
	(8.5)	(20.2)	(51.4)	(12.5)	(7.4)	(12.6)
Num Obs.	650	650	650	410	650	400
Bandwidth	5-Deg	5-Deg	5-Deg	3-Deg	5-Deg	MSERD
Mean PM10 South of Huai						
2003-2012	95	95	95	95	94	94
Specification	Linear	Quad	Cubic	Linear	Local Linear	Local Linear
FEs	Year	Year	Year	Year	Year	Year

Notes: The regression coefficients in this table are the estimated effects of the Huai River Policy on air pollution. The MSE-optimal RD bandwidth in Panel B, Column 6, is 2.88 degrees, using the methodology of Calonico et al. (2014). Year fixed effects are included in all regressions. Regressions and means in Columns 5 and 6 are weighted by an Epanechnikov kernel. Only monitors with at least ten years of data included in each regression. Standard errors are clustered by monitor.

*** p<0.01, ** p<0.05, * p<0.1

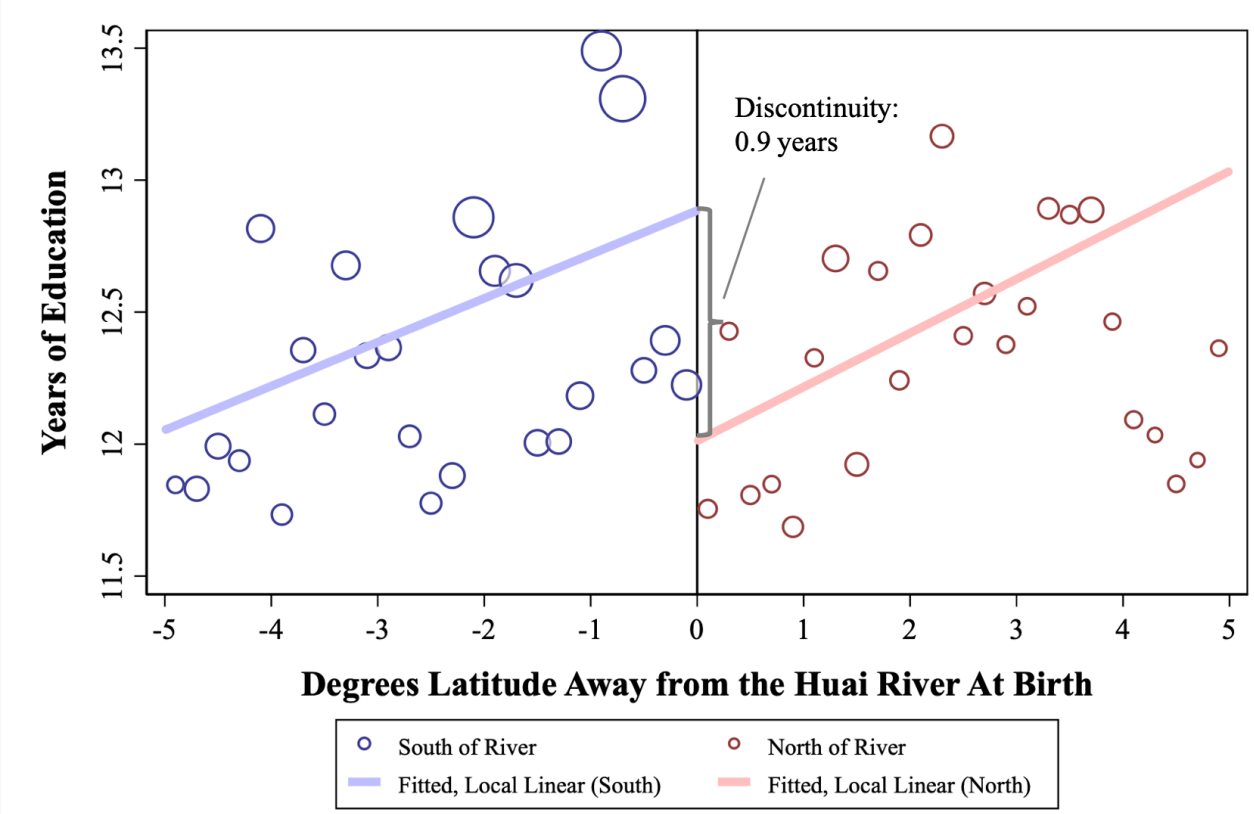
Figure 4: Pollution Discontinuity – Five-Degree Sample (Main)





Notes: Each bubble shows a monitor's mean pollution concentration, over 1981-2004 for TSP and over 2003-2012 for PM_{10}. For each data series, data are restricted to monitors with at least ten years of data. Lines estimated using local linear regression.

Figure 5: Education Discontinuity – No Fixed Effects



Notes: Each bubble shows the mean years of education within a 0.2-degree (22 km) latitude bandwidth relative to the Huai River, for those born during 1975-1982, who were ages 23-30 at the time of the 2005 census. The horizontal axis plots *Hukou* distance from the Huai River. Each bubble's size is proportional to its population, weighted by sampling weights. The lines show the local linear prediction using an Epanechnikov kernel and 5-degree bandwidth. The vertical distance at the Huai River (zero on the horizontal axis) between the purple local linear line and the red local linear line is the estimated discontinuity in years of education.

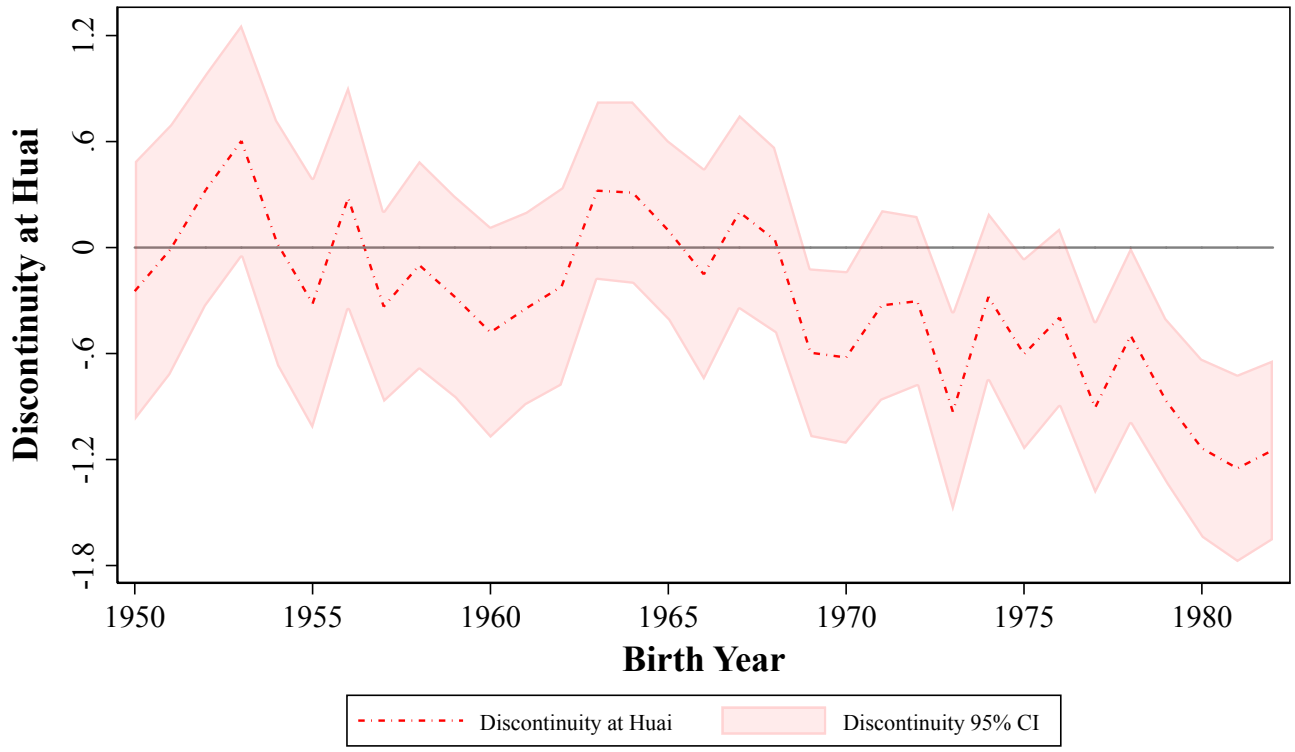
Table 4: Education Discontinuities – With Birth Year FEs

	Polynomial				Local Linear	
	(1)	(2)	(3)	(4)	(5)	(6)
All						
North of Huai	-0.66***	-1.17***	-0.87***	-1.03***	-0.83***	-1.03***
	(0.17)	(0.24)	(0.30)	(0.21)	(0.18)	(0.23)
Observations	38,885	38,885	38,885	26,932	38,885	24,718
Bandwidth (degrees)	5	5	5	3	5	2.632
Mean Education South of Huai	12.5	12.5	12.5	12.6	12.6	12.7
Females						
North of Huai	-0.58***	-1.20***	-0.81**	-1.06***	-0.79***	-1.00***
	(0.19)	(0.26)	(0.32)	(0.23)	(0.20)	(0.24)
Observations	20,766	20,766	20,766	14,437	20,766	12,941
Bandwidth (degrees)	5	5	5	3	5	2.587
Mean Education South of Huai	12.4	12.4	12.4	12.5	12.5	12.7
Males						
North of Huai	-0.75***	-1.12***	-0.96***	-0.99***	-0.87***	-1.06***
	(0.17)	(0.25)	(0.34)	(0.22)	(0.18)	(0.24)
Observations	18,119	18,119	18,119	12,495	18,119	11,687
Bandwidth (degrees)	5	5	5	3	5	2.709
Mean Education South of Huai	12.7	12.7	12.7	12.8	12.8	12.9
Specification	Linear	Quad	Cubic	Linear	Loc. Lin.	Loc. Lin.
Birth Year FEs	X	X	X	X	X	X

Notes: The regression coefficients in this table are the estimated effects of the Huai River Policy on years of education. All regressions are weighted by sampling weights. Regressions and means in columns 5 and 6 are furthermore weighted by an Epanechnikov kernel. The bandwidths in column 6 are the MSE-optimal bandwidths from Calonico et al. (2014). The sample consists of birth cohorts 1975-1982 with urban Hukou who are within the specified bandwidth from the Huai River. Due to differences in the estimated coefficients on the distance-to-Huai variables by regression, the all-gender estimate is not the weighted mean of the female estimate and the male estimate. Standard errors are clustered on Hukou county.

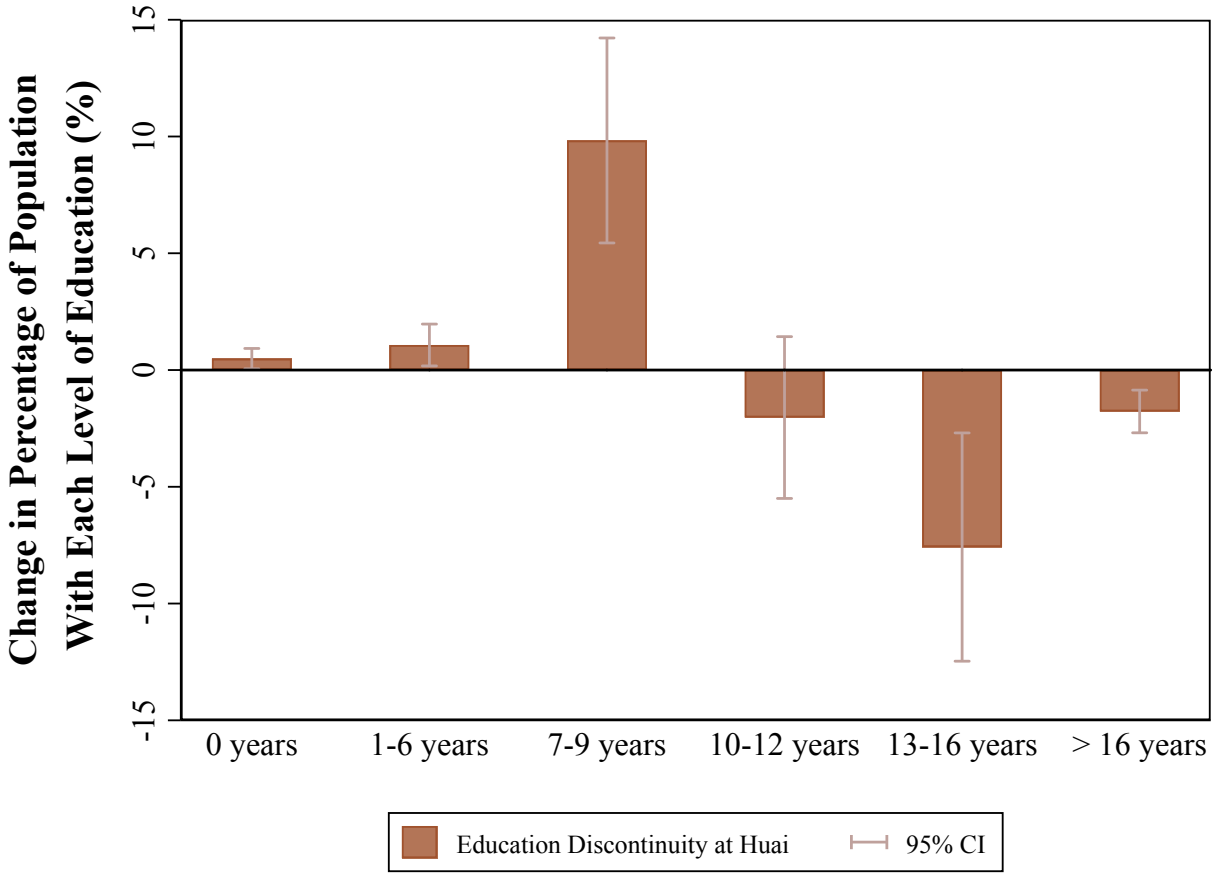
*** p<0.01, ** p<0.05, * p<0.1

Figure 6: Education Discontinuity By Cohort



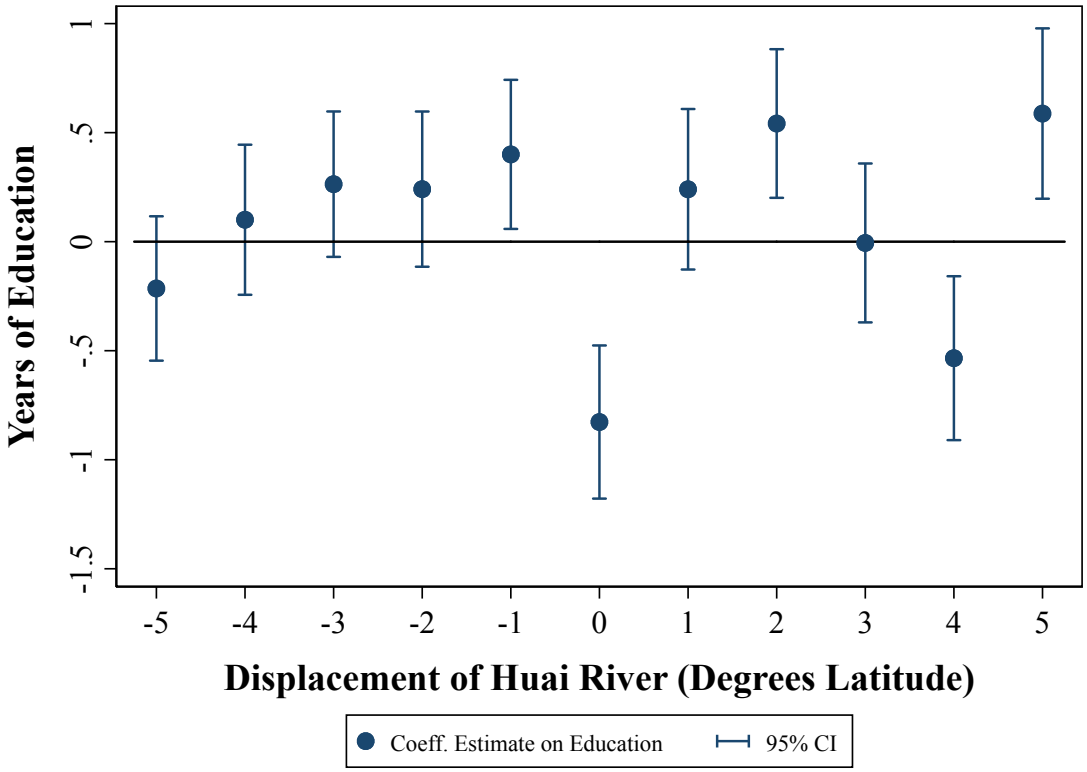
Notes: The discontinuities are estimated using an Epanechnikov kernel and five-degree bandwidth with separate regressions for each birth year. All regressions are weighted by sampling weights and have standard errors clustered at the *Hukou* county level.

Figure 7: Discrete Education Discontinuities



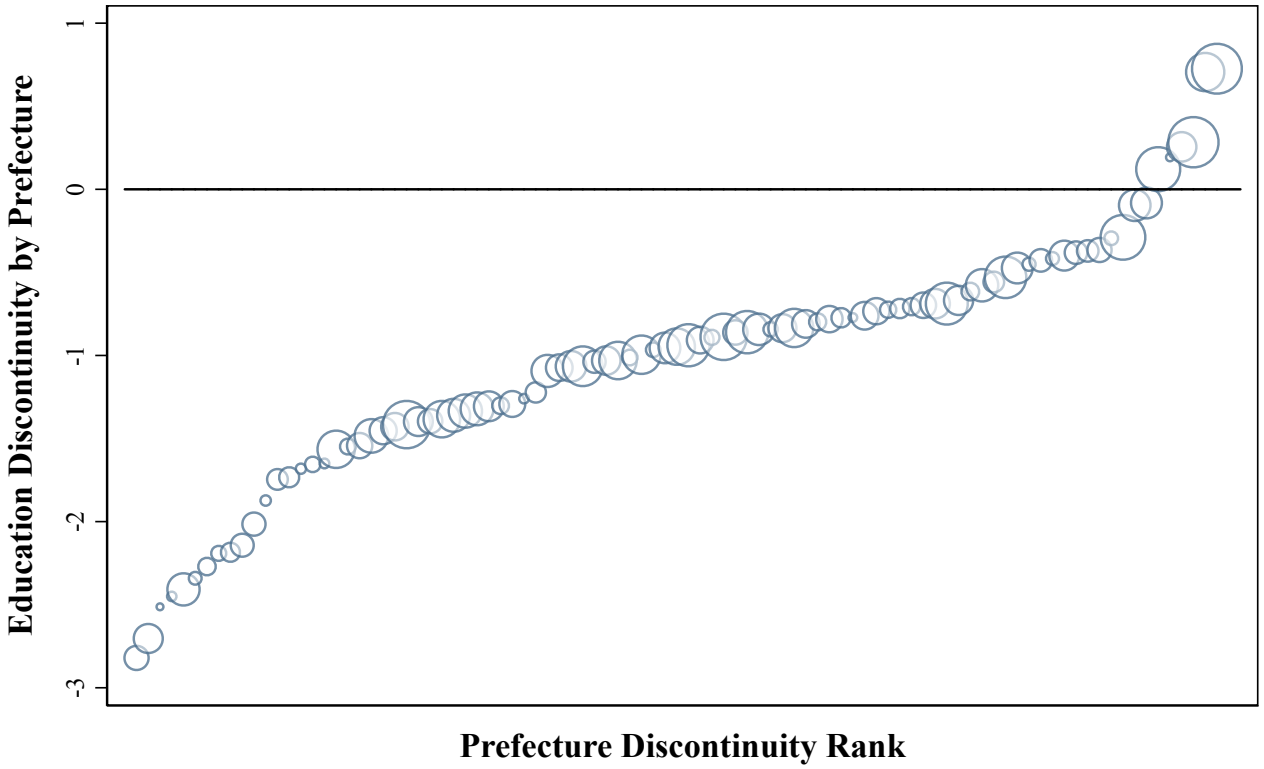
Notes: Each discontinuity estimate is from a separate local linear regression with an Epanechnikov kernel, sampling weights, and birth-year fixed effects. The sample consists of birth cohorts 1975-1982 with urban *Hukou* who are within a 5-degree bandwidth of the Huai River. The kernel-weighted mean proportions south of the Huai River, by education category, are 0.2% with 0 years, 2.1% with 1-6 years, 22.5% with 7-9 years, 33.9% with 10-12 years, 39.8% with 13-16 years, and 1.6% with more than 16 years.

Figure 8: Education Placebo Test – Shift Huai River



Notes: Each point in this figure is the estimated discontinuity in years of education using displaced Huai River boundaries at 1-degree latitude increments from the actual location of the river. Each discontinuity is estimated using a local linear regression with an Epanechnikov kernel, 5-degree bandwidth, birth-year fixed effects, and sampling weights. The sample consists of birth cohorts 1975-1982 with urban *Hukou*. Standard errors are clustered on *Hukou* county.

Figure 9: Education by Prefecture North of the Huai River



Notes: This figure shows the contribution of each prefecture to the discontinuity estimate of -0.83 years of education. Plotted are the prefectural means of residuals for people within 5-degree north of the Huai River, from a regression where years of education are regressed on distance to the river, the interaction of distance with a north-of-Huai indicator, and birth year fixed effects, and all these coefficients are constrained to have the same values as in the regression for Column 5 of Panel A of Table 4. The size of each prefecture's bubble is proportional to the population of the prefecture. Omitted from the graph are three prefectures where the number of observations is fewer than 10 (i.e. fewer than 5,000 urban residents). The regression uses local linear Epanechnikov kernel weights and sampling weights, and is restricted to the 1975-1982 urban *Hukou* birth cohorts. Standard errors are clustered on *Hukou* county.

Table 5: Income Discontinuities

	(1)	(2)	(3)	(4)	(5)	(6)
All						
North of Huai	-0.488***	-0.499***	-0.249***	-0.184***	-0.199***	-0.219***
	(0.050)	(0.053)	(0.062)	(0.035)	(0.036)	(0.043)
Observations	30,476	30,476	12,613	30,267	30,267	21,053
Bandwidth (degrees)	5	5	1.604	5	5	3
Mean Log Income South of Huai	6.92	6.96	7.07	7.26	7.27	7.30
Females						
North of Huai	-0.480***	-0.496***	-0.355***	-0.203***	-0.213***	-0.236***
	(0.056)	(0.060)	(0.072)	(0.051)	(0.053)	(0.065)
Observations	15,016	15,016	7,129	14,772	14,772	10,387
Bandwidth (degrees)	5	5	1.849	5	5	3
Mean Log Income South of Huai	6.84	6.88	6.98	7.15	7.17	7.19
Males						
North of Huai	-0.500***	-0.505***	-0.281***	-0.182***	-0.195***	-0.221***
	(0.049)	(0.052)	(0.066)	(0.046)	(0.048)	(0.059)
Observations	15,460	15,460	6,304	15,228	15,228	10,486
Bandwidth (degrees)	5	5	1.626	5	5	3
Mean Log Income South of Huai	7.00	7.04	7.15	7.36	7.38	7.40
Specification	Linear	Loc. Lin.	Loc. Lin.	Linear	Loc. Lin.	Loc. Lin.
Birth Year FEs	X	X	X	X	X	X
Residence County FEs				X	X	X

Notes: The regression coefficients in this table are the estimated effects of the Huai River Policy on monthly income (with zero income values coded as missing). All regressions are weighted by sampling weights. Regressions and means in columns 2, 3, 5, and 6 are furthermore weighted by an Epanechnikov kernel. The bandwidths in column 3 are the MSE-optimal bandwidths from Calonico et al. (2014). The sample consists of birth cohorts 1975-1982 with urban Hukou who are within the specified bandwidth from the Huai River. Note that in the regressions using residence county fixed effects, the identification comes from migrants, people whose residence counties differ from their *hukou* counties, and males are disproportionately migrants relative to females. Due to differences in the estimated coefficients on the distance-to-Huai variables by regression, the all-gender estimate is not the weighted mean of the female estimate and the male estimate. Standard errors are clustered on Hukou county.

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Migration Discontinuities – 1975-1982 Birth Cohorts

	Polynomial				Local Linear	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Migrated To New Prefecture						
North of Huai	-0.000	0.031*	0.016	0.026*	0.011	0.024
	(0.012)	(0.018)	(0.023)	(0.015)	(0.013)	(0.015)
Observations	38,885	38,885	38,885	26,932	38,885	26,247
Bandwidth (degrees)	5	5	5	3	5	2.911
Panel B: Migrated To New Province						
North of Huai	-0.004	0.028**	0.022	0.019	0.006	0.019
	(0.009)	(0.013)	(0.018)	(0.012)	(0.010)	(0.012)
Observations	38,885	38,885	38,885	26,932	38,885	23,697
Bandwidth (degrees)	5	5	5	3	5	2.521
Specification	Linear	Quad	Cubic	Linear	Loc. Lin.	Loc. Lin.
Quad Hukou Pref. Pop.	X	X	X	X	X	X
Birth Year FEs	X	X	X	X	X	X
Migration Rate South of Huai						
New Prefecture	0.13	0.13	0.13	0.12	0.12	0.11
New Province	0.10	0.10	0.10	0.09	0.09	0.08

Notes: The regression coefficients in this table are the estimated effects of the Huai River Policy on migration. A person is considered to have migrated if their prefecture (or province) of residence is different from their Hukou prefecture (or province). All regressions are weighted by sampling weights. Regressions and means in columns 5 and 6 are furthermore weighted by an Epanechnikov kernel. The bandwidth in column 6 is the MSE-optimal bandwidth from Calonico et al. (2014). The sample consists of birth cohorts 1975-1982 with urban Hukou who are within the specified bandwidth from the Huai River. Standard errors are clustered on Hukou county.

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Income Discontinuity – Heckman Selection Model

	(1) All	(2) Migrants	(3) Heckman	(4) Heckman	(5) Heckman	(6) Heckman	(7) Heckman	(8) Heckman	(9) Heckman
Main:									
North of Huai	-0.202***	-0.189***	-0.199***	-0.203***	-0.185***	-0.047	-0.209***	-0.191***	-0.121**
	(0.037)	(0.051)	(0.048)	(0.050)	(0.057)	(0.063)	(0.050)	(0.050)	(0.060)
Heckman First-Stage:									
North of Huai			0.263***	0.246***	0.224***	0.195***	0.252***	0.264***	0.035
			(0.069)	(0.068)	(0.067)	(0.066)	(0.068)	(0.069)	(0.070)
Hukou cty distance to Beijing			-0.010						0.429**
			(0.008)						(0.180)
Hukou cty distance to Dongguan				-0.055***					-2.391
				(0.012)					(1.479)
Hukou cty distance to Guangzhou					-0.076***				1.259***
					(0.014)				(0.434)
Hukou cty distance to Shanghai						0.020***			0.141***
						(0.003)			(0.026)
Hukou cty distance to Shenzhen							-0.048***		1.053
							(0.011)		(1.160)
Hukou cty distance to Tianjin								-0.000	-0.482***
								(0.007)	(0.156)
Inverse Mills:									
athrho			-0.088	-0.131	0.045	1.749***	-0.175	-0.011	1.636***
			(0.066)	(0.161)	(0.332)	(0.134)	(0.133)	(0.143)	(0.155)
lnsigma			-0.664***	-0.660***	-0.666***	-0.017	-0.655***	-0.667***	-0.081
			(0.015)	(0.022)	(0.018)	(0.052)	(0.024)	(0.014)	(0.061)
Birth Year FE	X	X	X	X	X	X	X	X	X
Residence County	X	X	X	X	X	X	X	X	X
Mean dist to dest			10.27	12.29	12.14	8.01	12.64	9.68	
Obs	28,306	4,756							
Obs main			4,756	4,756	4,756	4,756	4,756	4,756	4,756
Obs selection			30,476	30,476	30,476	30,476	30,476	30,476	30,476
Adj R^2	0.392	0.307							

Notes: The outcome variable in all regressions is log income. A person has migrated if their residence prefecture is different from their *Hukou* prefecture. All regressions are weighted by Epanechnikov kernel and sampling weights. The sample consists of birth cohorts 1975-1982 with urban *Hukou* who are within a 5-degree bandwidth of the Huai River. Standard errors are clustered on *Hukou* county. Distances are in 00s of km. Selection in the first-stage of the Heckman is for migration to any prefecture. For Heckman models, residence county fixed effects are only applied to the second-stage; moreover, the first-stage coefficients are probit coefficients.

Appendix

Table 8: 2005 Prefectural-level Education Expenditure (Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Education expenditure per pupil (¥)					Total expenditure per capita (¥)					Education share of total (0-1)				
North of Huai	-455.0	-2,477.0***	-2,514.3***	-2,549.5**	-2,264.0*	-518.0*	-1,763.6***	-1,765.0***	-1,893.8**	-2,029.5*	0.011	-0.007	-0.007	-0.013	-0.005
	(449.6)	(852.1)	(855.2)	(1,104.6)	(1,259.0)	(289.9)	(551.9)	(554.9)	(875.1)	(1,047.8)	(0.011)	(0.020)	(0.021)	(0.022)	(0.025)
Dist from Huai		462.1***	350.2	293.1	460.6		285.2***	280.7*	331.6	432.0		0.004	0.002	0.006	0.013
		(166.9)	(229.3)	(363.5)	(581.3)		(108.4)	(149.2)	(284.5)	(454.5)		(0.004)	(0.006)	(0.006)	(0.009)
North * Dist			239.1	365.5	-219.5			9.7	-10.7	-111.7			0.003	0.000	-0.021*
			(335.0)	(451.1)	(592.8)			(217.9)	(308.1)	(494.7)			(0.008)	(0.011)	(0.013)
Observations	141	141	141	141	92	142	142	142	142	93	141	141	141	141	92
Bandwidth (degrees)	5-Deg	5-Deg	5-Deg	5-Deg	3-Deg	5-Deg	5-Deg	5-Deg	5-Deg	3-Deg	5-Deg	5-Deg	5-Deg	5-Deg	3-Deg
Specification	Linear	Linear	Linear	Local Linear	Linear	Linear	Linear	Linear	Local Linear	Linear	Linear	Linear	Linear	Local Linear	Linear
Mean South of Huai	2,600.8	2,600.8	2,600.8	2,776.4	2,759.7	2,426.6	2,426.6	2,426.6	2,548.1	2,600.4	0.141	0.141	0.141	0.141	0.140

Table 9: 2005 Prefectural-level Education Expenditure (Excl Shanghai) (Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Education expenditure per pupil (¥)					Total expenditure per capita (¥)					Education share of total (0-1)				
North of Huai	-115.5	-1,566.3**	-1,620.8**	-1,538.4***	-1,078.8**	-259.7	-1,065.4**	-1,086.5**	-1,131.3**	-1,136.7*	0.010	-0.009	-0.009	-0.015	-0.009
	(372.0)	(714.8)	(709.8)	(484.1)	(429.3)	(217.1)	(418.8)	(418.3)	(474.6)	(579.3)	(0.011)	(0.021)	(0.021)	(0.023)	(0.025)
Dist from Huai		328.9**	93.4	-37.5	-21.6		183.0**	85.7	82.2	68.8		0.004	0.003	0.007	0.015
		(139.2)	(190.6)	(156.8)	(313.8)		(81.8)	(112.7)	(148.3)	(273.8)		(0.004)	(0.006)	(0.006)	(0.009)
North * Dist			495.8*	696.1**	262.7			204.7	238.6	251.5			0.002	-0.001	-0.023*
			(276.5)	(309.8)	(334.7)			(163.4)	(189.6)	(336.3)			(0.008)	(0.011)	(0.013)
Observations	140	140	140	140	91	141	141	141	141	92	140	140	140	140	91
Bandwidth (degrees)	5-Deg	5-Deg	5-Deg	5-Deg	3-Deg	5-Deg	5-Deg	5-Deg	5-Deg	3-Deg	5-Deg	5-Deg	5-Deg	5-Deg	3-Deg
Specification	Linear	Linear	Linear	Local Linear	Linear	Linear	Linear	Linear	Local Linear	Linear	Linear	Linear	Linear	Local Linear	Linear
Mean South of Huai	2,261.3	2,261.3	2,261.3	2,331.7	2,286.1	2,168.3	2,168.3	2,168.3	2,209.2	2,241.9	0.142	0.142	0.142	0.142	0.141

Table 10: Income Discontinuity – Heckman Selection Model (North Interaction in First-Stage)

	(1) All	(2) Migrants	(3) Heckman	(4) Heckman	(5) Heckman	(6) Heckman	(7) Heckman	(8) Heckman	(9) Heckman
Main:									
North of Huai	-0.202***	-0.189***	-0.206***	-0.196***	-0.039	-0.053	-0.204***	0.008	-0.124**
	(0.037)	(0.051)	(0.049)	(0.051)	(0.062)	(0.061)	(0.050)	(0.066)	(0.059)
Heckman First-Stage:									
North of Huai			0.081	-0.585**	-0.664***	0.366***	-0.459**	0.356***	-6.148**
			(0.169)	(0.233)	(0.223)	(0.074)	(0.233)	(0.116)	(2.586)
Hukou cty distance to Beijing			-0.023						-0.182
			(0.016)						(0.610)
North x Hukou cty distance to Beijing			0.020						0.785
			(0.017)						(0.692)
Hukou cty distance to Dongguan				-0.084***					-0.053
				(0.016)					(2.250)
North x Hukou cty distance to Dongguan				0.070***					-4.132
				(0.019)					(4.720)
Hukou cty distance to Guangzhou					-0.083***				0.930
					(0.016)				(0.648)
North x Hukou cty distance to Guangzhou					0.077***				-3.144
					(0.019)				(2.025)
Hukou cty distance to Shanghai						0.031***			0.159***
						(0.004)			(0.046)
North x Hukou cty distance to Shanghai						-0.032***			-0.421***
						(0.006)			(0.158)
Hukou cty distance to Shenzhen							-0.072***		-0.955
							(0.015)		(1.715)
North x Hukou cty distance to Shenzhen							0.059***		7.448*
							(0.019)		(4.508)
Hukou cty distance to Tianjin								0.009	-0.001
								(0.010)	(0.576)
North x Hukou cty distance to Tianjin								-0.012	-0.410
								(0.011)	(0.633)
Inverse Mills:									
athrho			-0.151*	-0.060	1.613***	1.722***	-0.142	1.793***	1.575***
			(0.082)	(0.201)	(0.196)	(0.133)	(0.154)	(0.151)	(0.155)
lnsigma			-0.658***	-0.666***	-0.073	-0.032	-0.659***	0.006	-0.111*
			(0.018)	(0.017)	(0.080)	(0.052)	(0.022)	(0.057)	(0.062)
Birth Year FE	X	X	X	X	X	X	X	X	X
Residence County	X	X	X	X	X	X	X	X	X
Mean dist to dest			10.27	12.29	12.14	8.01	12.64	9.68	
Obs	28,306	4,756							
Obs main			4,756	4,756	4,756	4,756	4,756	4,756	4,756
Obs selection			30,476	30,476	30,476	30,476	30,476	30,476	30,476
Adj R^2	0.392	0.307							

Notes: Sample consists of birth cohorts 1975-1982 with urban *Hukou* who are within a 5-degree bandwidth of the Huai River. Outcome variable in all regressions is log income. A person has migrated if their residence prefecture is different from their *Hukou* prefecture. All regressions are weighted by Epanechnikov kernel and sampling weights. Standard errors are clustered on *Hukou* county. Distances are in 00s of km. Selection in the first-stage of the Heckman is for migration to any prefecture. For Heckman models, residence county fixed effects are only applied to the second-stage; moreover, the first-stage coefficients are probit coefficients