Department of Economics, University of Warwick

Monash Business School, Monash University

as part of

Monash Warwick Alliance

Environmental Factors and Internal Migration in India

Kenji Komeda

Warwick-Monash Economics Student Papers

December 2021 No: 2021-20

ISSN 2754-3129 (Online)

https://warwick.ac.uk/fac/soc/economics/research/wmesp/

The Warwick Monash Economics Student Papers (WM-ESP) gather the best Undergraduate and Masters dissertations by Economics students from the University of Warwick and Monash University. This bi-annual paper series showcases research undertaken by our students on a varied range of topics. Papers range in length from 5,000 to 8,000 words depending on whether the student is an undergraduate or postgraduate, and the university they attend. The papers included in the series are carefully selected based on their quality and originality. WM-ESP aims to disseminate research in Economics as well as acknowledge the students for their exemplary work, contributing to the research environment in both departments.

"*We are very happy to introduce the Warwick Monash Economics Student Papers (WM-ESP). The Department of Economics of the University of Warwick and the Economics Department at Monash University are very proud of their long history of collaboration with international partner universities, and the Monash Warwick Alliance reflects the belief in both Universities that the future will rely on strong links between peer Universities, reflected in faculty, student, and research linkages. This paper series reflects the first step in allowing our Undergraduate, Honours, and Masters students to learn from and interact with peers within the Alliance.*"

Jeremy Smith (Head of the Department of Economics, University of Warwick) and Michael Ward (Head of the Department of Economics, Monash University)

Recommended citation: Komeda, K. (2021). Environmental Factors and Internal Migration in India. Warwick Monash Economics Student Papers 2021/20

[1] Warwick Economics would like to thank Lory Barile, Gianna Boero, and Caroline Elliott for their contributions towards the selection process.

Environmental Factors and Internal Migration in India

Kenji Komeda*

Abstract

This paper estimates the effect of air pollution, water pollution and water scarcity on internal migration in India using gravity model with 2SLS estimation. It contributes to the literature by first incorporating nationwide migrants and those three environmental factors into the analysis. The migration data is drawn from the Indian Census 2001 and 2011 and provides us with state-district pair-wise migration flows for certain time periods. With a wide range of data sources including Indian government platforms and satellite data, this study compiles a rich and comprehensive dataset. We find that the increase in air pollutant (PM2.5) at origin pushes out migrants, with larger influence on male than female. This paper also discovers, with more robust evidence, that the increase in groundwater level, a proxy for water scarcity level, at origin leads to less out-migrants and increase in groundwater at destination pulls more in-migrants for both genders. However, consistent evidence on water pollutants was not found.

JEL Classification: J16, J61, O15, Q25, Q53

Keywords: Internal Migration, Pollution, Water Scarcity, Gender Inequality, Gravity Model.

---

*kjkomeda@gmail.com

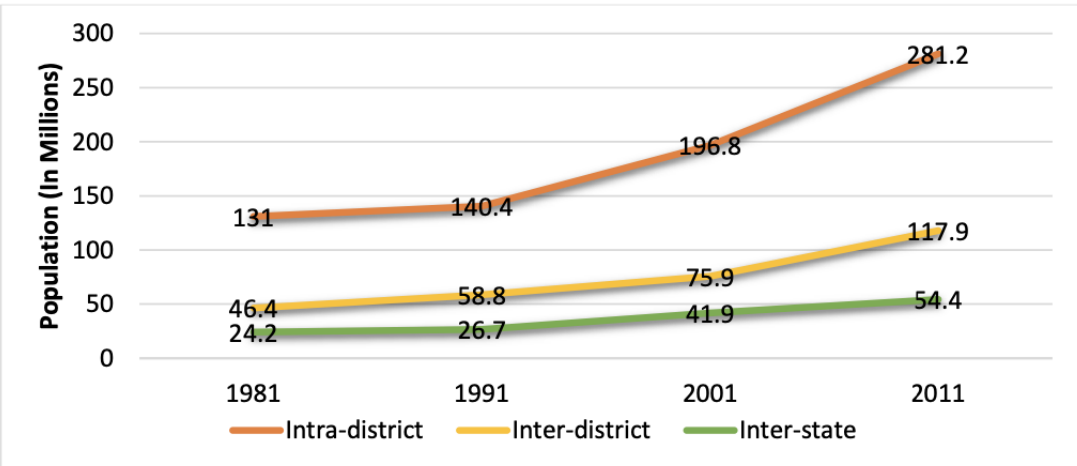
Appendix and do files are available:
https://www.dropbox.com/sh/2vjqdlvrbbv9u3u/AAAXjzNjO2tx4li1OH8FU66qa?dl=0

I am indebted to my supervisor Dr. Clement Imbert for his recommendation, feedback and guidance throughout this dissertation as well as for sharing migration data and relevant codes for my reference. I would also like to thank my wife and my family for their support.

1 Introduction

Internal migration has been becoming pervasive feature inherent in development in India, where unprecedented scale of urbanization in the next decade is expected (Prabha et al., 2020). Indeed, the Economic Survey (Government of India, 2017) enumerated that the annual internal migrants were 5-6 million on average. The main reasons of migration are employment and education for male and marriage and employment for female (Rajan and Neetha, 2018). Under the nationwide urbanization trends, primary origin locations tend to be densely populated but less urbanized states while destination sites are more industrialised. Figure 1 shows that shot-distance migrants indeed account for the majority. It stems from underlying social welfare scheme such as non-portability for subsidized grain, inflexible residential permission and insufficient formal insurance system hedging risks of migration (Kone et al., 2018; Munshi and Rosenzweig, 2016).

Figure 1: The Time Trends of Internal Migrants in India from 1981 to 2011



Source: Rajan and Bhagat (2021), created based on the Indian Census 1981-2011

With more active population mobility, the concept of environmental migration is driven by various forms of environmental and climatic stress and has brought another perspective to analyse the demographic dynamics (McLeman and Smit, 2006). Regarding environmental stress, the health impacts of air pollution and water pollution, in turn the values of those qualities in India have been recently elucidated (Shannon et al., 2018). Moreover, water scarcity is another crucial issue inducing mass migration in India, due to heavy reliance on water resources by agricultural sectors (Fishman et al., 2013).

Due to the direct connection with human health and security, those environmentally-induced issues have been highlighted lately (Neidell, 2004; Mannucci and Franchini. 2017; Sekhri, 2014; Sekhi and Hossain, 2020). Accordingly, a rich set of environmental regulations has been put into effect since 1970s, accompanied by government's continuous monitoring air pollutants, water pollutants as well as water scarcity (Greenhouse and Hanna, 2011; CGWB, 2019). Nonetheless, India has the highest number of pollution-linked deaths, which is approximately 2.3 million annually (Global Alliance on Health and Pollution, 2019). On the other hand, India faces crisis with water availability of around 1,100 cubic meter per person, below internationally recognised threshold of water stress of 1,700 and close to dangerous threshold of 1,000 (World Bank, 2019). In these terms, India provides a compelling setting to investigate the scale of environmental migration.

Yet, the impact of air pollution on internal migration in India has not been studied well. Existing studies mainly use China as a reference country, and the research context in most cases depends on China-specific policy background, which provides insufficient external validity. Moreover, there is a paucity of evidence about the effect of water pollution and water scarcity. The scope of previous research is also limited to specific regions or population groups. Furthermore, there is no study that incorporates all three environmental factors and examines relative effects on internal migration. Exploring the relative significance is relevant considering that those factors themselves are likely to be correlated. Additionally, understanding the comparative scales of impact on behavioural response has values in terms of policy formulation.

Intuitively, increase in pollutants should push population away or attract less migrants, while increases in water resources would pull more migrants or help to keep its population from leaving. The study uses Two Stage Least Square (2SLS) estimation using instrumental variables (IVs) to test the theoretical predictions. With a wide range of sources, we compile rich and comprehensive dataset with 2-year observations of migration flows. Air pollution, water pollution and water scarcity are measured by Particle Matter 2.5 (PM2.5), Biological Oxygen Demand (BOD) and Groundwater Level (GWL), respectively. We find that the increase in PM2.5 at origin leads to increase in out-migration, with larger influence on male than female. This paper also finds, with more robust evidence, that increase in GWL at origin leads to less out-migrants and increase in groundwater at destination pulls more in-migrants across both genders. The magnitude of impact is larger for water scarcity than air pollution. The results are consistent even in a range of validity checks.

This paper is structured as follows; Section 2 reviews the findings of relevant literature, which subsequently identifies the gaps in-between. Section 3 describes empirical strategy. Section 4 presents overview of our datasets followed by data description. Section 5 provides preliminary analysis on the data. Section 6 reports our findings along with robustness checks. Section 7 identifies research limitations and concludes.

2 Literature

2.1 Climatic Variability and Migration

The impacts of climate variability, in the broad sense of environmental factors, have been popularised in light of the growing risk of climate change in the present era. Previous studies show that weather abnormalities such as rising temperature and higher frequency of drought, push factors, are associated with both internal and international migration in India (Dallmann and Millock, 2016; Sedova and Kalkuhl, 2020). Viswanathan and Kumar (2015) finds that especially weather-related decline in the value of agricultural output leads to increase in out-migration rate. Sedova and Kalkuhl (2020) accordingly indicate those climate migrants tend to be lower-skilled labours largely depending on their incomes from agricultural production. Moreover, farmers are likely to migrate when there is insufficient availability of instructions, knowledge and technology on agriculture to adapt to new climate conditions (Jha et al., 2018). The recent study also investigates relationships between migration and existences of nature or heatwave as environmental amenities and disamenities (Winkler and Rouleau, 2020).

2.2 Air and Water Pollutions and Internal Migration

With increasing severity of pollution, health risks have been studied as factors of migration. Chen et al. (2017) first studies effects of air pollution on migration. They find that deteriorated air quality from 1960 to 2010 is responsible for the population decline of 5% due to reduced inflows and increased outflows. Such negative effects on both population attraction and individuals' willingness to settle down the region are also documented in the subsequent studies (Li et al., 2017; Cui et al., 2019; Li et al., 2020; Liu and Yu, 2020; Germani, 2021). Other studies elucidate labour sorting effects of air pollution. Using Chinese firm-level financial data, Xue et al. (2020) finds that polluted areas experience falls in executive talent and high-skilled labours. Khanna et al. (2021) extends the research and finds that skilled labour had higher sensitivity, being likely to emigrate in response to pollution, than unskilled ones. They thus indicate that pollution induce inefficient spatial distribution of skills across cities, leading to aggregate productivity loss. This result attributes to underlying health impacts of pollutions. Exposure to air pollutants has been linked to premature death, increased mortality, respiratory disorder and asthma, in turn, lower performance of labour and increased hospitalisation (Chay and Greenstone, 2003; Deryugina et al., 2019; He et al., 2016; Archsmith et al., 2018; He et al., 2019). Similarly, daily exposure to polluted water could increase digestive cancer rate, increased diarrhea-related deaths and higher likelihood to become disabled (Ebenstein, 2012; Garg et al., 2018; Lai, 2017).

2.3 Water scarcity and migration

The variability in water access stemming from increasing stress on hydrological cycle is a new dimension of recent studies. Previous literature reveals that water scarcity is associated with a variety of negative social and economic consequences such as heightened risks of riot and conflict and lower agricultural output (Almer et al., 2017; Domania, 2020; Vallino et al., 2020). India is the largest consumer of groundwater in the world (Fishman et al., 2013). Therefore, Indian subcontinent, experiencing increased climate variability, has been particularly exposed to such risks due to water depletion (World Bank, 2016). Sekhri (2014) examines the impact of access to groundwater on poverty and discovers that groundwater depth below a certain cutoff level leads to higher rural poverty rate. Similarly, Sekhri and Hossain (2020) proves that groundwater scarcity heightens the risk of sexual violence against women due to their longer time spent outside to collect water. Regarding the impact on agricultural output, Blakeslee et al. (2020) shows that water scarcity simply results in falls in farm income and wealth. Consistent with their findings, Zeveri et al. (2020) evaluates the effect of water access on short-term migration in rural India and finds that less stable water access leads to higher likelihood for family to send migrants to secure alternative income. Fishman et al. (2013) further specialises on the social grouping of the migrants in depth. The survey indicates that groundwater depletion encourages younger males to migrate to urban centres. Moreover, it reveals much lower rate of migration who belong to scheduled castes with little land ownership.

2.4 Gaps in Literature

While these findings offer valuable implications, there are still gaps to be filled. Regarding the impact of air pollution on migration, most of the previous research presented above are conducted using China as a sample context. Thus, their external validity is a concern in terms of policy formulation. This dissertation will address the need to provide another country-specific result by using India as a sample. Also, to the best of our knowledge, the relationship between water pollution and migration has not been explored. Therefore, our study will provide the first evidence on the field. When it comes to water scarcity, Fishman et al. (2013) and Zeveri et al. (2020) provide the effect on migration. However, the former research only focuses on a specific part of India, Rural Northern Gujarat; thus, it does not provide a nationwide evidence. The latter uses short-term migration samples, which does not fully capture the impact of water depletion on the people's eventual choice of residence. By using long-term migration dataset, this dissertation gives more holistic estimates of the impact of water scarcity on internal migration. Finally, our study incorporates three factors, air pollution, water pollution and water scarcity, to estimate overall impacts on internal migration and to analyse interrelation among the factors.

3 Empirical Specification and Method

3.1 Theoretical Framework and Econometric Specification

Gravity model is commonly-used paradigm for migration analysis because of its consistency with theories (Poot et al., 2016; Imbert and Papp, 2020). In line with the previous literature, our study will use a typical gravity migration model to examine the environmental determinants of bilateral migration flows, i.e. 'push' and 'pull' factors:

$$ln\ M_{ij,t} = D_{j,t} - O_{i,t} - C_{ij,t} + \varepsilon_{ij,t} \tag{1}$$

where $M_{ij,t}$ is the number of bilateral migrants from origin state i to destination district j at time t,[1] $O_{i,t}$ ($D_{j,t}$) is time-varying district characteristics of origin (destination) including populations, and $C_{ij,t}$ is migration cost between the two places at time t including physical distances. As developed by Beine and Parsons (2012), the logarithm of emigration rate, defined as the ratio of out-migrants over population in the location at a given time period, is frequently used as dependent variable. However, with such design, we cannot identify whether changes in the emigration rate stem from changes in migration flow or population (Sousa, 2013). Hence, our study uses the migration flows as dependent variable while controlling for population dynamics.

By expanding the equation (1), we will first estimate a Pooled Ordinary Least Square (POLS) specification:

$$\ln M_{ij,t} = \alpha_0 + \alpha_1 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} Envi_{i,k} + \alpha_2 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} Envi_{j,k} + \alpha_3 \ln O_{i,t-1} + \alpha_4 \ln D_{j,t-1}$$
$$\alpha_5 \ln P_{ij} + \delta_t + u_{ij,t} \qquad (2)$$

The dependent variable measures migration flows from state i to district j between year $t-1$ and t and $u_{ij,t}$ is error term. The principal variables of interest, $\frac{1}{\theta}\sum_{k=t-\theta}^{t-1} Envi_{i,k}$ ($\frac{1}{\theta}\sum_{k=t-\theta}^{t-1} Envi_{j,k}$) represent the mean measures of air pollution, water pollution and/or groundwater level in origin state i (destination district j) between year $t-\theta$ and $t-1$, where $\theta > 1$. Given maximum data availability, we set $\theta = 3$ in our analysis. Hence, the main variables of interest are 3-year average values. The specification is designed to capture the accumulated effects of environmental factors prior to migration since we expect reaction of individuals to perceived changes in the environment would be slow and also the changes need to be observed for certain period of time. By focusing on migration flows within the last 1-year duration, we maintain a strict causality between the push and/or pull factors and the decision to migrate. Our hypotheses are that higher (lower) pollutants in air and water and lower (higher) groundwater level works as push (pull) factors on migration.

[1] Our data have information on the number of migrants from each origin state to destination district.

A set of origin (destination) control variables $O_{i,t-1}$ $(D_{j,t-1})$ includes variables explaining origin (destination) characteristics, in addition to total populations. In developing countries, adverse climate disasters such as drought and floods mostly force individuals to escape since vulnerables could be significantly affected. Particularly in India, 58.2% (234.1 million) and 54.6% (263.1 million) of population respectively in 2001 (2011) are engaged in agriculture (Census 2001 and Census 2011). In such environment, weather factors directly affect their well-being; hence act as a strong incentive to migrate. Indeed, Burgess (2017) finds that hot days are associated with sharp decline in agricultural output and rural wage accordingly as well as sharp increase in mortality rates. Also, Sedova and Kalkuhl (2020) finds that temperature and precipitation abnormalities work as push factors on inter-state migration in India. Hence, we include mean temperature and rainfall indicator. The ratio of total crime rates against women ($Cr_{j,t-1}/Cr_{i,t-1}$) is used as an alternative indicator for regional safety level as religious or caste-based violence in particular motivating migration (Mitra and Raj, 2014; Sousa, 2013).

Following common practice, migration cost is captured by distance between state i and district j (d_{ij}) and language proximity (l_{ij}) (Bodvarsson and Van den Berg, 2009; Imbert and Papp, 2017). We also control for caste similarity between two places expressed by the ratios of scheduled castes (SC) and scheduled tribes (ST) in the destination compared to those of origin. In India, these two are important social factors as potential determinants of migration (Mitra and Murayama, 2008).[2] Due to the discrimination, they are likely to be day labourers or casual workers, presumably the most vulnerable groups to pollutions (Dubey et al., 2006). Anderson (2011) shows that they have less access to water resources such as groundwater in times of drought. Hence, it can be hypothesised that they would prefer to stay within their communities or be unable to migrate to other areas, and that they only migrate to the place where they can find similar communities. Indeed, Bhattacharya (2002) finds that SC incidence in rural areas is associated with lower outmigration rates.

Income or Gross Domestic Product per capita is often included in gravity model. However, Dallmann and Millock (2016), in their climate analysis on internal migration, excludes it from their specifications since income is endogenous to climate variables. This follows

[2] In 2001 (2011), 16.2% (16.6%) of population belonged to SC, and 8.2% (8.6%) to ST in 2001 (2011), based on the caste, ethnic and religious categorization of the society by ‘Hindu Varna’ System. (Chandramouli, 2013)

the argument by Dell et al. (2014) on the trade-off between omitted variable bias and over-controlling problem that arises by including regressors endogenous to the main variable of interests. Including such variables biases the estimation of the net effect of main variables on migration. For this reason, we cannot include income since it is endogenous to pollutions and access to groundwater level (Khanna et al., 2021; Sekhri, 2014; Blakeslee et al., 2020; Zeveri et al., 2020).

Adding period fixed-effects (FE) specification improves simple POLS estimation as it captures time-specific unobserved heterogeneity such as global macroeconomic trends and national-level policies. Origin, destination and state-district pair-wise FE are also considered. However, adding such FE drops our main variables based on our gravity construction. Hence, given data limitation, we use only period FE (δ_t) for our baseline specification.[3] Still in robustness checks, we estimate specifications with those FEs. To adjust for autocorrelation within state-district pairs over time, we use robust clustered standard errors at pair-wise level.

3.2 Poisson Pseudo Maximum Likelihood Estimation

POLS estimation is feasible but will yield inconsistent estimates. This is because we estimate gravity models and migration flows with zero values are excluded from its estimation, leading to potential selection bias. Instead, we use Poisson Pseudo-Maximum Likelihood (PPML) estimator with robust clustered standard errors, that is robust to the presence to zeros and even to heteroscedasticity caused by the zeros, even in a non-linear model (Silva and Tenreyro, 2006).[4]

3.3 Identification Problems

Identification means that the variables of interest are independent of the error term conditional on all control variables. However, both POLS and PPML specifications may still produce biased estimates due to the failure of this assumption. Primary issue is endogeneity of main variables of interest such as omitted variable bias, simultaneity and

[3] The details explained in the following data section of instrumental variables.
[4] The PPML estimator has been performed well in trade literature and applied to migration literature as well (Flowerdew and Aitkin, 1982; Imbert and Papp, 2020).

measurement error. Specification (3) may suffer from omitted variable bias through time-variant unobserved heterogeneity. Indeed, regional GDP is not controlled due to data limitation. More generally, estimates are susceptible to unmeasured confounding factors which jointly determine pollutions (water scarcity) and migration. For instance, the underlying economic structure is prone to be positively associated with pollutions while regions with more active manufacturing plants attract migrants (Borjas, 1999; Clark et al., 2007). With regard to simultaneity, it is not clear whether pollutions (water scarcity) affect migration or migration affects pollutions (water scarcity) via demographic change. Another issue is measurement error, which if present, causes correlation between independent variables and error term, resulting in bias towards zero (Pischke, 2007). Our variables respectively measure each environmental issue from one dimension. That means they are just proxies for the risks of air pollution, water pollution and water scarcity, latter of which, rather than the former, affects migration.

3.4 Instrumental Variables Approach

To get around these potential problems, earlier papers maintain exogeneity by 2SLS approach (Sousa, 2013; Khannna et al., 2021). The paper follows the procedures using different types of instruments for each environmental factor:

$$\ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} Envi_{i,k} = \gamma_0 + \gamma_1 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} IV_{i,k} + \gamma_2 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} IV_{j,k} + \gamma_3 \ln O_{i,t-1} + \gamma_4 \ln D_{j,t-1}$$
$$\gamma_5 \ln P_{ij} + \delta_t + \epsilon_{ij,t} \qquad (3)$$

$$\ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} Envi_{j,k} = \rho_0 + \rho_1 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} IV_{i,k} + \rho_2 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} IV_{j,k} + \rho_3 \ln O_{i,t-1} + \rho_4 \ln D_{j,t-1}$$
$$\rho_5 \ln P_{ij} + \delta_t + \vartheta_{ij,t} \qquad (4)$$

$$\ln M_{ij,t} = \beta_0 + \beta_1 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} Envi_{i,k} + \beta_2 \ln \frac{1}{\theta} \sum_{k=t-\theta}^{t-1} Envi_{j,k} + \beta_3 \ln O_{i,t-1} + \beta_4 \ln D_{j,t-1}$$
$$\beta_5 \ln P_{ij} + \delta_t + \zeta_{ij,t} \qquad (5)$$

where $\frac{1}{\theta}\sum_{k=t-\theta}^{t-1} IV_{i,k}$ $(\frac{1}{\theta}\sum_{k=t-\theta}^{t-1} IV_{j,k})$ is instrumental variables for origin (destination) and $\epsilon_{ij,t}$, $\vartheta_{ij,t}$ and $\zeta_{ij,t}$ are error terms. Equation (3) and (4) are the first stages for origin and destination and equation (5) is the second stage specification of the 2SLS system respectively. The right-hand-side of variable in equation (3) and (4) measures the average of environmental value for a given year periods for either origin or destination. We instrument environmental value for origin and destination with a relevant instrument respectively, conditional on controls and period FE.

In line with the previous attempt (Chen et al., 2017; Khanna et al, 2021), air pollution is instrumented by thermal inversion, which is a meteorological phenomenon where the temperature of the atmospheric layer in higher altitude becomes higher than that in lower altitude. Since temperature is normally lower in higher altitude and air moves from hot to cold places, the concentration of pollutants on the ground naturally decreases. However, thermal inversion hinders vertical circulation of air and traps polluted air near the ground, leading to higher concentration. We use the total number of thermal inversions happened in the last given years in each area. The valid instruments have to meet relevance and exclusion restrictions, respectively meaning that there is correlation with endogenous variables conditional on covariates and no direct effect on outcome variable other than through endogenous variables conditional on covariates (Angist and Pischke, 2008). The relevance condition here specifies the existence of effect of thermal inversion on air pollution. This can be checked through F-statistic. However, one concern is that thermal inversion can coincide with other weather patterns on the ground and the weather variations themselves can affect migration, violating exclusion restriction. In order to deal with the confounding issue, available weather variables are included in our specifications for identification. There is still a chance that thermal inversions can coincide with macroeconomic trends, but period FE will deal with the issue.

There is no well-established instrument for water pollution in relation to migration. However, as Ebenstein et al. (2015) employs, we use average rainfall, conditional on weather controls. Since our study focuses on organic pollutants within water resources, rainfall is expected to reduce intensity of pollutants in a given volume. Same concerns and approach as thermal inversions are applied to this instrument. For groundwater level, Ryan and Sudarshan (2020) uses a variety of instruments such as rock types, aquifer type and underground fractures. Among which, our study uses the probability of igneous rocks primarily occupying the geological composition of the area given data availability.

Igneous rock is formed through the process of cooling and solidification of lava or magma. Compared with other main rocks i.e. sedimentary and metamorphic rocks, igneous rock is hard with high density and not porous; hence, groundwater can sit on the top and fill up the reservoir with more ease. We expect that the probability of rock types will not directly nor indirectly affect migration flows.

4 Data and Variable Construction

4.1 State-District Pair-wise Migration Data

Data on state-district pair-wise migration flows is drawn from the Indian Census 2001 and 2011, conducted by the Office of the Registrar General & Census Commissioner, India, decennially since 1871. In the Census, migration flows are defined by the last residence and current residence with different durations of stay at the current place. The advantage of the Census is that it records spatial distinction of urban and rural areas of migration flows (i.e. urban-urban, urban-rural, rural-rural and rural-urban migration).[5] Although the Census has limitation for showing only last residence in state-level while the current residence is in district-level. However, to the best of our knowledge, it has the most comprehensive spatial coverage in India. Regarding time-dimension, we focus on migration within the last 1 year duration in order to maintain strict causality in individuals' response to push and pull factors and to minimise the measurement error related to the subsequent migration. Hence, our analysis uses the panel of pairs of origin-destination migration flows with 4 patterns of urban-rural combinations within the last 1 year from each Census 2001 and 2011. In line with our econometric framework, our dependent variable is thus the gross bilateral migration flow from origin state to destination district between year $t-1$ to year t. The administrative division of districts has been changed during the observation periods, but we use the 2001 boundaries to construct a panel.

[5] The area is assigned as urban if it satisfies either of following conditions: existence of municipality, companies, accommodations and notified area committee; minimum population of 5,000, more than 75% of male working population engages in non-agricultural sectors, population density with more than 400 persons per square km (India Census 2011).

4.2 Measures of State & District-Level Environmental Factors

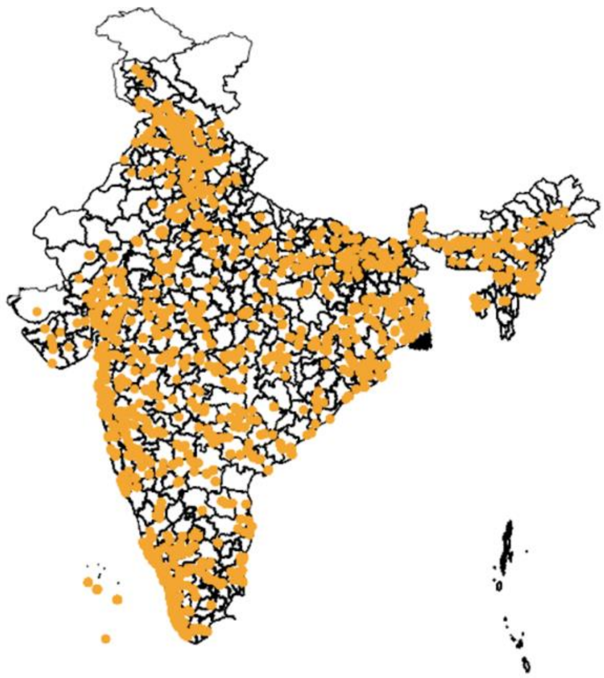
Previous studies on India's air pollution primarily use the particle matter with diameter less than $10\mu m$ (PM10), Sulfur Dioxide (SO2) and Nitrogen Dioxide (NO2) data from the Central Pollution Control Board (CPCB) (Greenstone and Hanna, 2014). It collects daily readings on air pollutants using a growing network of monitoring stations across the country, under the National Air Quality Monitoring Programme (NAMP) from 1987 to present. However, since the programme has been gradually developed through the decades, the stations are not equally located in geographical terms. Hence, there are, for example, missing observations for small-sized districts at specific time periods. In order to make the best of our comprehensive migration data, we instead measure state-level and district-level air quality using Global Annual PM2.5 Grids satellite data that can cover all regions of India (Freeman et al., 2019). Van Donkelaar et al. (2016) first combines Aerosol Optical Depth (AOD) retrievals from the MODIS, MISR and SeaWiFS, and subsequently predicts grid cell global ground-based observations of PM2.5 using Geographically Weighted Regression (GWR). The calibrated estimates represent annual concentration of PM2.5 without dust and sea-salt, with high grid cell resolution of 0.01 degrees, from 1998 to present. Moreover, the data is advantageous in that PM2.5 is considered as the best indicator of air pollution in terms of health risks. Indeed, fine particles (diameter<2.5 μm) are more hazardous than larger particles (2.5 μm <diameter<10 μm) on mortality, cardiovascular and respiratory systems (WHO, 2014). By matching each grid to the 2001 Indian district boundaries, we construct annual average PM2.5 panel for each state and district.[6]

The CPCB also administers water quality monitoring stations under the National Water Monitoring Programme (NWMP). It covers rivers and tributaries, creeks, lakes and ponds, drains and canals, and groundwater resources. We retrieve observations between the year 1986 and 2011. Although the programme is still under development and the data have not collected for some districts, it contains the richest information on water quality in India. We match latitude and longitude information for each monitoring station with the geospatial boundaries of the states and districts in 2001, in order to construct a panel of annual mean water quality. Figure 2 illustrates the locations of the water quality monitoring stations in 2011. Although the CPCB collects 28 annual-level measures of water quality, previous studies use organic pollutants i.e. Biochemical Oxygen Demand

[6] The shapefile of India recorded for Census 2001 and 2011 are used for the geospatial matching.

(BOD), Dissolved Oxygen and Faecal Coliforms due to its most consistent observations throughout the years. Among which, we use BOD as it is one of the measurements with consistently significant presence in CPCB Official Water Quality Criteria (CPCB, 2013). BOD indicates the quantity of oxygen necessary for decomposition of organic waste in water. Although it cannot capture the chemical pollutants, it is considered as one of the most common measure of organic pollutants (Greenstone and Hanna, 2011). Using organic pollutants helps us to isolate itself from chemical air pollutants, avoiding potential multicollinearity problem.

Figure 2: Geospatial Distribution of Water Quality Monitoring Stations



Source: CPCB, 2021

As a proxy for water scarcity, our study uses groundwater level (GWL). Indeed, a typical form of water stress stems from the depletion of groundwater resources since India is the largest consumer of groundwater (World Bank, 2010). The data is retrieved from India Water Resources Information System. The database reserves district-wise annual average depth of groundwater (in metres below ground level) from 1993, constructed from individual observations of approximately 16,000 monitoring wells across India. It is collected by the Indian Central Groundwater Board.

4.3 Controls and State & District-level Characteristics

Weather can work as a significant confounding factor since it may affect both migration and our main variables of interest (Dallmann and Millock, 2016; Panda et al., 2012; Khanna et al., 2020; Ebenstein, 2012; Viswanathan and Kumar, 2015). Due to India's large geographical territory, we can observe high climate variability across districts (Pant et al., 2016). In previous literature, weather variables are used as indicators of drought or excess precipitation and as key factors of migration decision (Dallmann and Millock, 2016; Sedova and Kalkuhl, 2020). In other cases, variables derived from temperature and rainfall are employed as instrumental variables for air and water quality respectively (Arceo et al., 2016; Chen et al., 2017; Khanna et al., 2020; Ebenstein, 2012).

Hence, we include mean temperature and rainfall variables as weather controls. We obtain those from ERA5-Land monthly averaged data from 1981, which is the reanalysis products of the European Centre for Medium-Range Weather Forecasts (ECMWF). It combines numerical weather prediction models with satellite-based and ground-based observations and constructs a globally comprehensive dataset. ERA5-Land has higher resolution grid of 0.1 degrees in latitude and longitude, which can provide measurements for all districts in India. We calculate state-level and district-level annual average temperature. In order to accurately capture geographic-specific weather effects, with rainfall data, we calculate the Standardised Precipitation Index (SPI) which is a standardised measure of drought on a range of timescales, invented by McKee et al. (1993). The raw long-term rainfall data is fitted to a gamma distribution, and the mean and standard deviation are calculated for past periods (i.e. 1981-2011). Based on that, SPI quantifies observed rainfall as the number of standard deviations above or below from the long-term mean. The advantage of SPI is that it explains the deviation for a period at a certain place as well as it makes the values of different locations comparable. It allows us to determine the length and intensity of drought. Thus, they are better indicators of weather as determinants of migration than absolute values of rainfall (Seiler et al., 2002). Following Dallmann and Millock (2016), we construct state and district-level variables for the annual magnitude (sum of SPI) of drought or excess precipitation.

Since migration decision may be complex, our study also controls for other social and cultural factors as proxies of migration costs. First is the index of language proximity. The Census of India provides district-level data on the number of people having each language as mother tongue. Using the data, we construct a variable representing the probability of randomly-chosen two individuals each from origin and from destination speak the same language. Second, the ratios of the number of SC and ST in destination over that of origin is included as a network factor. Since these two measurements are obtained from the Census, we have information for each urban and rural area of districts. Third is crime rate against women from 2001 to 2011 constructed from Crime in India issued by the National Crime Records Bureau. It provides the number of women-related crime in district-level.

Finally, population and distance are also included as typical gravity variables. Origin and destination population can represent the scale of economies as push and pull gravity. Since the migration data do not contain actual physical distance for each migration, we create a proxy for the travel distance. We first calculate geodesic distances between the geographical centroids of each pair of districts.[7] In order to transform them into the distances between the pairs of state and district, we compute population weighted average distance from districts of a given state to a given destination district.[8]

4.4 Instrumental Variables

Thermal inversion is constructed from the MERRA-2 of the National Aeronautics and Space Administration (NASA). In particular, we utilise the product M216NPANA version 5.12.4, with the spatial resolution of 0.5 degree × 0.625 degree grid. It records 6-hour average air temperature at 42 layers from land surface to 36,000 meters. First, the grid data are aggregated to each state and district. Next, temperature difference between the third layer (540 meters) minus the second layer (320 meters) is calculated within each 6-hour period.[9] The positive difference indicates the existence of thermal inversion. Finally, we count the total number of incidents when the region experiences thermal inversion across 6-hour lapses within 1-year period. The data on average rainfall is from

---

[7] The geodesic distance is the shortest path between two points on a sphere or ellipsoid.

[8] $State\ to\ District\ Distance = \sum_{i=1}^{n} \frac{DistrictPopulation_{it} * Distance_{ij}}{StatePopulation_{lt}}$, where l is the origin state, i is a district in the origin state, n is the number of districts in an origin state, and j is the destination district.

[9] Chen et al. (2017) and Khanna et al. (2021) uses the difference between the second and first layer. However, the first layer has too many missing observations across India, leading to omitting the most of our samples. We therefore use the third and second layers. Chen et al. (2017) checked the validity and consistency of using the third layer to compute thermal inversion.

ERA5-Land. We take the annual mean of monthly rainfall for each year in a given state or district.

Geological data is obtained from an open government platform, Bhukosh – the Geological Survey of India. It records lithological compositions such as types of rocks and soils, with the geographical coverage for each lithology across India. We extract the primary geological characteristics for each recorded location and categorise into 4 types, i.e. igneous, metamorphic, sedimentary and other rocks. By aggregating across each state and district, the probability of igneous rocks mainly forming the regional ground is calculated. One disadvantage is that rock type is geographical-specific and not time-varying. It hence limits the inclusion of state or destination FE into our baseline model since it simply leads to the drop of main variable(s).

4.5 Merged Dataset

India has been splitting its districts during the observation periods – 593 districts in 2001 and 640 districts in 2011. In line with our interest to precisely investigate the change in patterns of migration flow, we make a panel which is defined by the district boundaries in 2001. However, some of our data sources provide information based on 2011 boundaries. In such cases, we first construct datasets following 2011 boundaries. Second, state and district names are cleaned in order to prevent from making duplicates or losing information by mistakes when merging. Third, based on the data types, the means or sums of variables are taken by collapsing with state and district names of 2001 boundaries. Fourth, cleaned datasets is merged with cleaned migration data by names of place and area types – urban, rural and total, separately for origin state and destination district. Due to inability to assign area-specific values, we use aggregate values of environmental and weather variables for each area types.[10] We repeat this process for 2001 migration data and for other datasets defined by 2001 boundaries. Finally, the 2011 dataset is appended to 2001 datasets. The final dataset consists of a balanced panel of 20,755 pairs of migration flows for each observed year of 2001 and 2011. However, due to the existence of missing observations in each variable, there is some extent of variability in the number of samples used in each econometric specification.

[10] This limits the accuracy of our research, but the analysis using aggregate values for each area type still gives the consistent results with our main analysis and have some policy implications (Table 26 in Appendix).

4.6 Descriptive Statistics

Table 1 illustrates summary statistics on the variables used in our analysis, with mean, standard deviation, the minimum and maximum values. Based on our econometric framework and data availability, the 3-year average value of the year 1998-2000 and 2008-2010 is used for environmental and weather variables. Crime variable is computed based on 2001 and 2010. For other variables, values in 2001 and 2011 are used. The mean of female migrants is higher than that of male migrants although standard deviation is also higher, indicating the variation around the mean is large. Since there are many zero values across migration flows of origin-destination pairs, log-transformation leads to fewer observations. Nonetheless, log-transformed number of migrants have much lower standard deviations because log operator reduces the effect of outliers. BOD has comparatively higher standard deviation than PM2.5 and GWL. Observations for thermal inversions are limited due to data availability. Due to the unbalanced missing observations across variables, we use different maximum number of samples available for each analysis.

Table 1: Descriptive Statistics

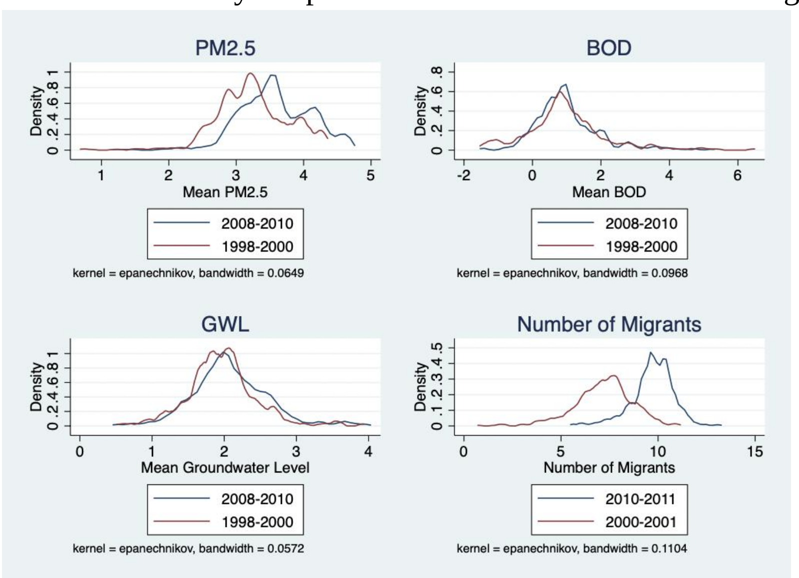
Variable	Count	Mean	Std. Dev.	Min	Max
Number of Migrants					
Total Migrants	41,510	445.84	4,474.42	0	377,092
ln(TotalMigrants)	27,633	3.06	2.25	0	12.84
Male Migrants	41,510	203.84	2,112.58	0	192,985
ln(MaleMigrant)	25,776	2.72	2.12	0	12.17
Female Migrants	41,510	242.00	2,402.23	0	184,107
ln(FemaleMigrant)	23,791	2.59	2.14	0	12.12
Main Variables					
ln(O_PM2.5)	41,510	3.24	0.56	1.30	4.63
ln(D_PM2.5)	41,510	3.41	0.56	0.68	4.70
ln(O_BOD)	36,766	1.17	1.07	-1.17	4.30
ln(D_BOD)	21,210	1	1.05	-1.54	6.50
ln(O_GWL)	32,615	2.02	0.45	0.91	3.09
ln(D_GWL)	38,570	2.03	0.51	0.50	3.98
Instrumental Variables					
ln(O_thermal inversion)	26,092	7.26	1.59	4.08	9.75
ln(D_thermal inversion)	22,190	4.98	1.31	-0.69	6.32
ln(O_mean rainfall)	39,138	-5.49	0.54	-6.69	-4.36
ln(D_mean rainfall)	40,740	-5.70	0.55	-7.56	-4.01
O_igneous rocks ratio	41,510	0.20	0.23	0	1
D_igneous rocks ratio	41,230	0.20	0.28	0	1
Controls					
ln(O_population)	41,510	15.97	2.12	11.01	19.11
ln(D_population)	41,510	14.08	1.04	10.34	16.77
ln(weighted distance)	41,510	13.86	0.71	0	15.38
language proximity	41,510	0.15	0.27	0	0.97
ln(SC ratio)	41,510	-0.88	5.75	-17.54	14.96
ln(ST ratio)	41,510	-2.18	6.40	-16.55	14.85
ln(O_crime)	40,917	6.89	2.41	0.51	10.06
ln(D_crime)	40,285	5.04	1.34	-0.41	8.51
ln(O_temperature)	39,138	3.08	0.31	2.09	3.33
ln(D_temperature)	40,460	3.14	0.34	-0.28	3.34
O_SPI	41,510	4.41	62.28	-117.87	258.19
D_SPI	41,510	0.26	3.61	-17.82	15.13

Notes: The Table presents the summary of the variables used in the analysis. The controls include typical gravity variables, social variables and weather variables. All explanatory variables, except for the ones expressed as probabilities (igneous rocks ratio and language proximity) and the SPI with negative values, are log-transformed for gravity equation.

5 Descriptive Patterns of Migration and Environmental Factors

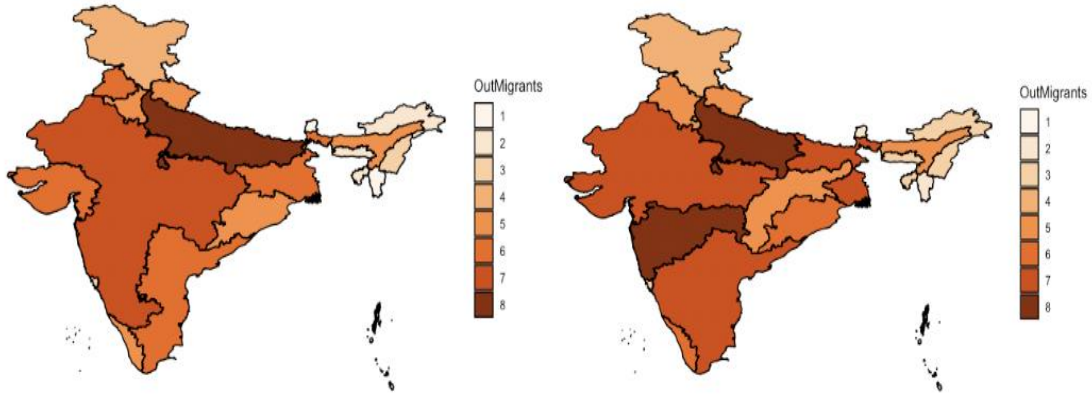
The Figure 3 provides kernel density estimates of 3-year average environmental factors for 1998-2000 and 2008-2010 and that of migrants for 2000-2001 and 2010-2011. It captures the spatial variations across Indian districts. The entire distributions of PM2.5 has shifted to the right over the decade, indicating worse air condition across country. The distributions for BOD and GWL have not changed meaningfully over the observation period. Contrarily, there has been a striking increase in the mean with narrower distribution for the number of migrants over the decade, suggesting migration becoming more pervasive. Correspondingly, the Figure 4 and 5 indicates the dispersed overall increase in both out-migrants and in-migrants across India. However, population movement still tends to concentrate on regions around industrial cities such as Delhi (North) and Mumbai (South West). Internal migration appears not very populous in the top North and East side. In conjunction with the Figure 6 showing geographical patterns of environmental factors, spatial correlation between pollutions and migrations are observed. The spatial indication for GWL is not clear, but GWL in most migrated areas appears to be lower.

Figure 3: Kernel Density Graphs of Environmental Factors and Migrants



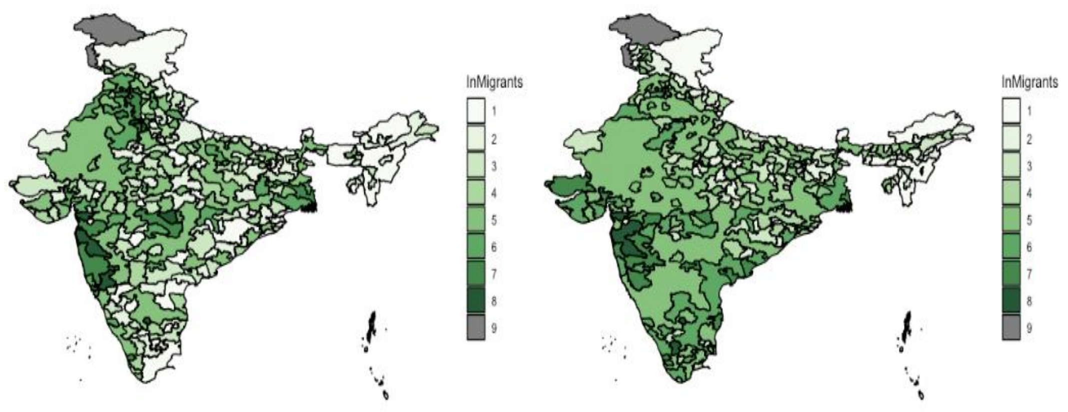
Notes: The figures provide the distribution of log-transformed 3-year average environmental factors (1998-2000 and 2008-2010) and the distribution of 1-year number of migrants (2001 and 2011) of our sample. An Epanechnikov kernel function is used for the construction.

Figure 4: The Geographic Distribution of the Share of Out-Migrants
for 2000-2001 and 2010-2011



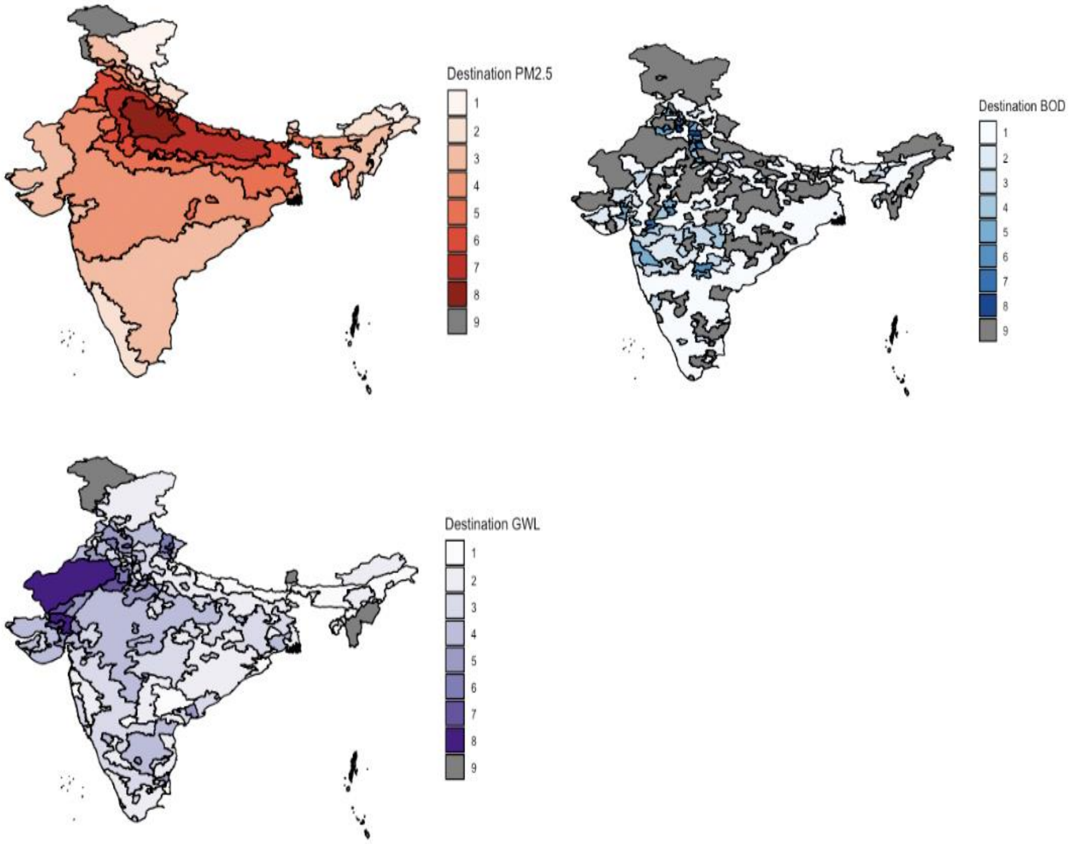
Notes: The out-migrant shares are ratio of out-migrants flowing out from the region over all out-migrants across India. Each label stands for each range of percentages; 1: 0-0.05%, 2: 0.05-0.1%, 3: 0.1-0.5%, 4: 0.5-1%, 5: 1-2.5%, 6: 2.5-5%, 7: 5-10%, 8: 10-20%, 9: more than 20%.

Figure 5: The Geographic Distribution of the Share of In-Migrants
for 2000-2001 and 2010-2011



Notes: The in-migrant shares are ratio of in-migrants flowing in to the region over all in-migrants across India. 1: 0-0.25%, 2: 0.25-0.5%, 3: 0.5-0.75%, 4: 0.75-1%, 5: 1-2.5%, 6: 2.5-5%, 7: 5-10%, 8: more than 10%, 9: insufficient data.

Figure 6: The Geographic Distribution of PM2.5, BOD and GWL for 2008-2010

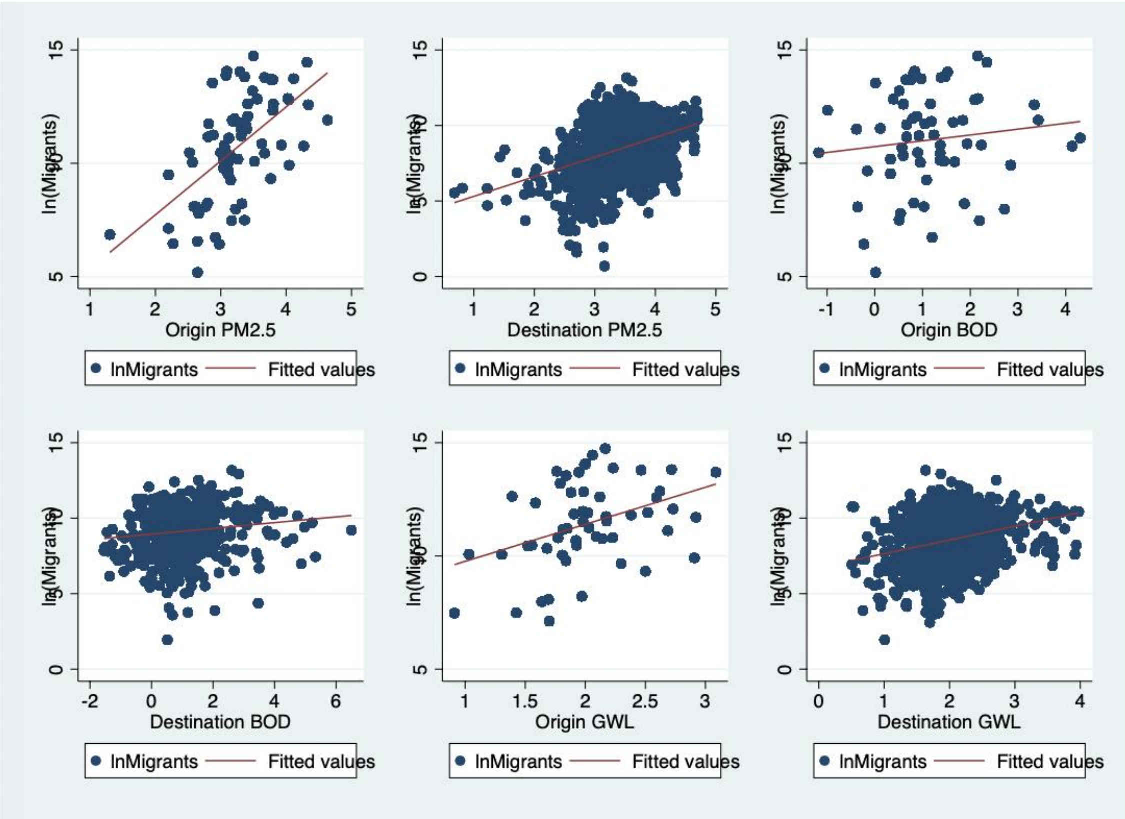


Notes: Each label stand for each range of concentration levels. PM2.5 labels: 1: 0-10, 2:10-20, 3: 20-30, 4: 30-40, 5: 40-50, 6: 50-60, 7: 60-80, 8: more than 80, 9: insufficient data, with unit $\mu g/m^3$. BOD labels: 1: 0-4, 2: 4-6, 3: 6-8, 4: 8-10, 5: 10-20, 6: 20-40, 7: 40-60, 8: more than 60, 9: insufficient data, with unit mg/L. GWL labels: 1: 0-5, 2: 5-7.5, 3: 7.5-10, 4: 10-15, 5: 15-20, 6: 20-25, 7: 25-30, 8: more than 30, 9: insufficient data, with unit m.

The Figure 7 shows scatter plots and fitted lines between each environmental factor and migrants. Although there is a certain degree of variation, all factors at origin and destination interestingly seem to have positive relationships with migrations. It is understandable in a sense. For pollutions, more polluted cities tend to have larger population with higher absolute number of out-migrants, while more polluted destination cities with more prosperous economic activities attract people. For GWL, origin cities with more GWL may have fewer population with less economic activities motivating residents to move out, while destination cities with more water resources attract people.

However, it is just observational relationships without controlling confounding factors and we thus find real associations and causal relationship in the following section.

Figure 7: The Relationship between Log-Migrants and Environmental Factors



Notes: The Figure plots relationships between three environmental factors and migration at origin and destination. All values are log-transformed.

6 Empirical Results

6.1 Main Results

Table 2 reports the results of different specifications of the gravity models laid out in the equations (2) for the effect of air pollution on migration. The POLS estimation, with weather and social controls as well as period FE (column 1), suggests significant correlations between PM2.5 and migration flows. It shows that, with statistical significance, 1% increase in PM2.5 at origin pushes out about 0.36% more migrants to other places on average. On the other hand, the same amount of increase of PM2.5 in destination attracts about 0.92% less migrants to the place. The PPML estimator can take account of the presence of zero values and indeed allows us to include more samples in regressions. It still presents the same sign as POLS on both coefficients with greater magnitude. Turning to Column 3 to 6, we consider the heterogeneity by gender. In general, the estimates indicate the similar effect on both gender category, although the significance reduces or disappears on the push effects for female. Given marriage is the primary reason for female while being work for male, this suggests a migration pattern that female is less responsive to air pollution since they need to stay their original place as long as they are single. Contrarily, the greater economic magnitude for male in both coefficients also indicates that male is more sensitive to pollution and flexible to move.

We should interpret those results with cautions. The estimates may suffer from endogeneity concern and may be corelated with unobserved time-varying determinants of migration within origin and destination, such as regional GDP, job opportunities and infrastructure, leading to omitted variable bias. Those potential omitted factors are likely to attract migrants, in turn, the estimates may be biased downwards for origin and upwards for destination. Also, positive out-migration flows may lower pollution level while positive in-migration may force it up. Therefore, such reverse causality may cause downward bias and upward bias for the estimates of origin and destination respectively.

Table 3 reports the results for the 2SLS estimations. As presented, the F statistics for the first stage regression is 540.525 for both genders, which are clearly over a recommended threshold of 10, indicating the relevance of instrument. The results indeed show a clear positive association between the frequency of thermal inversion and air pollution as expected. We find that air pollution at origin has positive and statistically significant effect on out-migration at 1 percent level. Also, consistent with the above bias discussion, the magnitude of the effect of air pollution on out-migration from origin is larger for total and male migrants. Specifically, 1% increase in PM2.5 induces an increase in out-migration by about 0.5% for both genders. It is consistent with previous findings (Chen et al., 2017). In contrast, the coefficient for destination is not statistically significant nor interpretable with large standard error. Surprisingly, the signs of coefficients for female are both flipped with less significance. In conjunction with the former argument, this may suggest that female, with less attention to air pollution, primarily migrates when they marry with the ones living near or the centre of urban area, who is likely to have higher income.

Table 2: The Effect of Air Pollution on Migration

	(1)	(2)	(3)	(4)	(5)	(6)
	Both genders		Male		Female	
DV: ln(Migrants)	POLS	PPML	POLS	PPML	POLS	PPML
ln(O_PM2.5)	0.359***	0.418**	0.442***	0.598***	0.172**	0.193
	(0.0649)	(0.175)	(0.0645)	(0.166)	(0.0674)	(0.192)
ln(D_PM2.5)	-0.921***	-0.970***	-0.930***	-1.081***	-0.798***	-0.835***
	(0.0501)	(0.138)	(0.0510)	(0.145)	(0.0503)	(0.139)
ln(O_population)	0.931***	0.0636	0.931***	-0.0586	0.700***	0.203
	(0.0458)	(0.128)	(0.0461)	(0.123)	(0.0500)	(0.137)
ln(D_population)	0.497***	1.101***	0.500***	1.284***	0.456***	0.929***
	(0.0437)	(0.293)	(0.0435)	(0.329)	(0.0466)	(0.266)
ln(weighted distance)	-1.792***	-1.844***	-1.617***	-1.634***	-1.859***	-2.029***
	(0.0366)	(0.0780)	(0.0362)	(0.0840)	(0.0385)	(0.0770)
language proximity	0.774***	1.444***	0.700***	1.379***	0.872***	1.530***
	(0.0873)	(0.136)	(0.0853)	(0.144)	(0.0903)	(0.137)
ln(SC_ratio)	0.0558***	-0.435***	0.0542***	-0.529***	0.0314**	-0.349***
	(0.0115)	(0.102)	(0.0117)	(0.111)	(0.0130)	(0.0981)
ln(ST_ratio)	-0.0485***	-0.0528***	-0.0476***	-0.0634***	-0.0494***	-0.0419***
	(0.00380)	(0.0112)	(0.00380)	(0.0119)	(0.00390)	(0.0112)
ln(O_crime)	-0.118***	0.196***	-0.166***	0.125	-0.0311	0.256***
	(0.0394)	(0.0743)	(0.0396)	(0.0765)	(0.0410)	(0.0775)
ln(D_crime)	0.279***	0.264***	0.219***	0.287***	0.294***	0.245***
	(0.0287)	(0.0989)	(0.0284)	(0.109)	(0.0300)	(0.0926)
ln(O_temperature)	0.0981	2.812***	0.195	3.034***	0.294*	2.744***
	(0.154)	(0.830)	(0.152)	(0.852)	(0.172)	(0.869)
ln(D_temperature)	-1.500***	0.0245	-1.310***	0.771	-1.455***	-0.689
	(0.257)	(0.783)	(0.260)	(0.829)	(0.261)	(0.784)
O_SPI	-0.000211	0.00241***	7.83e-06	0.00265***	-0.000382*	0.00222***
	(0.000189)	(0.000756)	(0.000190)	(0.000746)	(0.000195)	(0.000783)
D_SPI	0.0230***	0.0252	0.0250***	0.0353*	0.0196***	0.0163
	(0.00467)	(0.0169)	(0.00467)	(0.0181)	(0.00494)	(0.0164)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	No	No	No	No	No	No
N	10,977	13,596	10,447	13,596	9,993	13,596
R^2	0.526		0.493		0.507	

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

Table 3: The Effect of Air Pollution on Migration (2SLS Estimation)

	(1)	(2)	(3)
First Stage	Both genders	Male	Female
DV: ln(O_PM2.5)			
ln(O_thermal inversion)	0.141***	0.144***	0.146***
	(0.0021)	(0.0021)	(0.0021)
DV: ln(D_PM2.5)			
ln(D_thermal inversion)	0.083***	0.0810***	0.0831***
	(0.003)	(0.003)	(0.0030)
F statistic	540.525	493.660	497.460
Second Stage			
DV: ln(Migrants)	Both genders	Male	Female
ln(O_PM2.5)	0.500***	0.753***	-0.0308
	(0.119)	(0.114)	(0.121)
ln(D_PM2.5)	-0.0518	-0.244	0.336**
	(0.161)	(0.165)	(0.171)
ln(O_population)	0.881***	0.839***	0.722***
	(0.0529)	(0.0542)	(0.0583)
ln(D_population)	0.330***	0.362***	0.264***
	(0.0527)	(0.0521)	(0.0563)
ln(weighted distance)	-1.791***	-1.611***	-1.867***
	(0.0376)	(0.0371)	(0.0397)
language proximity	0.328***	0.271**	0.430***
	(0.115)	(0.114)	(0.120)
ln(SC ratio)	0.0408***	0.0465***	-0.000578
	(0.0129)	(0.0130)	(0.0145)
ln(ST ratio)	-0.0342***	-0.0415***	-0.0221***
	(0.00533)	(0.00529)	(0.00571)
ln(O_crime)	-0.0800*	-0.102**	-0.0406
	(0.0446)	(0.0443)	(0.0466)
ln(D_crime)	0.331***	0.265***	0.347***
	(0.0304)	(0.0302)	(0.0319)
ln(O_temperature)	-0.110	-0.00788	0.120
	(0.162)	(0.158)	(0.182)
ln(D_temperature)	-0.979***	-0.838***	-0.785**
	(0.306)	(0.310)	(0.316)
O_SPI	-2.88e-05	0.000564**	-0.000870***
	(0.000257)	(0.000255)	(0.000270)
D_SPI	0.0562***	0.0507***	0.0640***
	(0.00764)	(0.00771)	(0.00825)
Year FE	Yes	Yes	Yes
Instrument	Yes	Yes	Yes
N	10,977	10,447	9,993
R^2	0.494	0.466	0.459

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

Table 4 provides the estimates of the effect of water pollution on migration.[11] As the naïve PPML estimates present, we find negative associations between origin water pollution and out-migration on one hand, and positive associations between destination water pollution and in-migration on the other hand. At a glance, this result appears to confirm the counterintuitive correlation that water pollutant gathers population. However, we here suspect reverse causality that increase in population causes increase in water pollution as well. After all, robust 2SLS estimations return intuitive signs for destination coefficients, suggesting 1% increase in water pollutant in destination leads to about 0.29% fall in in-migration (Table 5).[12] However, the signs for origin coefficients remain negative. Furthermore, with an alternative measure of water pollution with Faecal Coliform[13], the overall results turned opposite.[14] The inconsistency may indicate identification failure by the IV, although the signs of coefficients for rainfall are negative with significance as expected, suggesting instrument relevance. For instance, although we control for SPI to capture rain-related factors to affect migration, it might not capture all weather variations nor rainfall-related outcomes. Therefore, the mean rainfall as an IV itself may still affect migration via unobserved omitted variables such as humidity but also economic factors like agricultural yield. It leads to a violation of exclusion restriction where the chosen IV becomes invalid. Given the fact that there is no well-established IVs in the context of water pollution and migration, our results also suggest the necessity of further investigation for causal inference.

[11] See Table 11 in Appendix for full estimates.
[12] See Table 12 in Appendix for full estimates.
[13] Faecal Coliform is another primary indicator for organic pollutants such as animal and human waste, which is strongly correlated with the presence of harmful pathogens in water (Greenhouse and Hanna, 2011)
[14] See Table 13 in Appendix for full estimates.

Table 4: The Effect of Water Pollution on Migration

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	Both genders		Male		Female	
ln(Migrants)	POLS	PPML	POLS	PPML	POLS	PPML
ln(O_BOD)	0.0259	-0.146***	0.00868	-0.0963*	0.0390**	-0.187***
	(0.0169)	(0.0530)	(0.0167)	(0.0535)	(0.0175)	(0.0541)
ln(D_BOD)	0.123***	0.170***	0.110***	0.206***	0.126***	0.140***
	(0.0152)	(0.0402)	(0.0154)	(0.0408)	(0.0157)	(0.0406)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	No	No	No	No	No	No
N	13,626	17,294	12,917	17,294	12,290	17,294
R^2	0.526		0.498		0.494	

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

Table 5: The Effect of Water Pollution on Migration (2SLS Estimation)

	(1)	(2)	(3)
First Stage			
	Both genders	Male	Female
DV: ln(O_BOD)			
ln(O_mean rainfall)	-0.801***	-0.818***	-0.842***
	(0.0306)	(0.0317)	(0.0325)
DV: ln(D_BOD)			
ln(D_mean rainfall)	-0.554***	-0.567***	-0.559***
	(0.0241)	(0.0248)	(0.0250)
F statistic	395.565	394.613	369.206
Second Stage			
DV: ln(Migrants)	Both genders	Male	Female
ln(O_BOD)	-0.202***	-0.198***	-0.230***
	(0.0556)	(0.0536)	(0.0534)
ln(D_BOD)	-0.288***	-0.311***	-0.155**
	(0.0669)	(0.0648)	(0.0669)
Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Instrument	Yes	Yes	Yes
N	13,626	12,917	12,290
R^2	0.483	0.449	0.462

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

Table 6 summarises the estimated results of the effect of groundwater level on migration.[15] All specifications show consistent signs. An increase in groundwater level of origin is associated with a fall in out-migrants while an increase in destination is associated with an increase in in-migrants. However, again there are endogeneity concerns. In this case, omitted variables such as aforementioned economic indicators can be negatively and positively correlated with out-migrants and in-migrants, respectively. Therefore, given the possible negative relationship between those economic indicators and GWL, the estimates may suffer from upward bias for origin and downward bias for destination. The presence of reverse causality is also suspected. Positive out-migration flows may help wells to recharge water while positive in-migration flows lead to faster consumption of stored water. Such reverse causality may cause upward bias and downward bias for origin and destination respectively. Consistent with the expected bias, the IV estimates of the effect of GWL on out-migration have lower coefficients and those on in-migration have larger coefficients for all columns in Table 7.[16] The instrument relevance is also as expected with significance. Compared with the former pollutants, the effect is economically more meaningful and there is less heterogeneity by gender. This suggests that groundwater is vital resources for livelihood and acts as stronger incentive for individuals' choice of migration.

Table 6: The Effect of Groundwater Level on Migration

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	Both genders		Male		Female	
ln(Migrants)	POLS	PPML	POLS	PPML	POLS	PPML
ln(O_GWL)	-0.0965***	-0.354***	-0.120***	-0.318***	-0.153***	-0.390***
	(0.0334)	(0.103)	(0.0328)	(0.101)	(0.0346)	(0.109)
ln(D_GWL)	0.346***	0.0440	0.321***	0.0950	0.314***	0.00521
	(0.0254)	(0.0608)	(0.0250)	(0.0652)	(0.0267)	(0.0601)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	No	No	No	No	No	No
N	22,368	28,577	21,136	28,577	19,805	28,577
R^2	0.490		0.462		0.464	

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

[15] See Table 14 in Appendix for full estimates.

[16] See Table 15 in Appendix for full estimates.

Table 7: The Effect of Groundwater Level on Migration (2SLS Estimation)

	(1)	(2)	(3)
First Stage			
	Both genders	Male	Female
DV: ln(O_GWL)			
ln(O_igneous rocks)	0.443***	0.437***	0.447***
	(0.0123)	(0.0128)	(0.0131)
DV: ln(D_GWL)			
ln(D_igneous rocks)	0.367***	0.365***	0.353***
	(0.0129)	(0.0131)	(0.0136)
F statistic	382.001	359.990	324.781
Second Stage			
DV: ln(Migrants)	Both genders	Male	Female
ln(O_GWL)	-0.560***	-0.565***	-0.474***
	(0.172)	(0.176)	(0.183)
ln(D_GWL)	0.814***	0.904***	0.860***
	(0.153)	(0.153)	(0.165)
Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Instrument	Yes	Yes	Yes
N	22,368	21,136	19,805
R^2	0.467	0.430	0.437

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

The Table 8 and 9 present estimation results with two or all three environmental factors incorporated in each specification.[17] Our sample is much smaller than previous discussion with higher standard errors due to missing observations. Nevertheless, POLS and PPML estimators show similar effects as we have used each single environmental factor as main variables shown in the above. This suggests that each factor independently affects migration flows. For 2SLS, given the suspect of invalidity for water pollution instrument, we gradually expand our specifications. Column 1 with air pollution and groundwater level as our variables of interest indicates that our findings still hold in terms of effect direction. Noticeably, the coefficients for GWL at destination are positive and statistically significant for all gender categories. It suggests the robust pull effects. However, once water pollutions are added to the specifications, the results diverge from our findings in the sense of economic significance and signs. This also supports our suspicion on the existence of underlying biases.

[17] See Table 16 and 17 for full estimates.

Table 8: The Effect of Three Environmental Factors on Migration

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	Both genders		Male		Female	
ln(Migrants)	POLS	PPML	POLS	PPML	POLS	PPML
ln(O_PM2.5)	0.113	0.123	0.253***	0.385*	-0.123	-0.220
	(0.0902)	(0.245)	(0.0895)	(0.232)	(0.0921)	(0.264)
ln(D_PM2.5)	-1.056***	-1.101***	-1.019***	-1.180***	-0.890***	-1.000***
	(0.0721)	(0.191)	(0.0719)	(0.205)	(0.0724)	(0.185)
ln(O_BOD)	-0.0308	0.178**	-0.0572**	0.138	0.0151	0.230**
	(0.0262)	(0.0898)	(0.0260)	(0.0909)	(0.0266)	(0.0907)
ln(D_BOD)	0.249***	0.254***	0.231***	0.305***	0.232***	0.206***
	(0.0240)	(0.0640)	(0.0242)	(0.0672)	(0.0245)	(0.0621)
ln(O_GWL)	-0.273***	-0.208*	-0.248***	-0.244*	-0.344***	-0.160
	(0.0600)	(0.123)	(0.0590)	(0.125)	(0.0613)	(0.129)
ln(D_GWL)	0.327***	-0.0236	0.273***	-0.0673	0.346***	0.00589
	(0.0494)	(0.116)	(0.0494)	(0.124)	(0.0512)	(0.110)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	No	No	No	No	No	No
N	6,121	6,956	5,898	6,956	5,731	6,956
R^2	0.572		0.540		0.556	

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

Table 9: The Effect of Three Environmental Factors on Migration (2SLS Estimation)

	(1)	(2)	(3)	(4)	(5)	(6)
Second Stage						
DV: ln(Migrants)	Both genders		Male		Female	
ln(O_PM2.5)	0.138	0.155	0.486*	0.522**	-0.265	-0.270
	(0.258)	(0.240)	(0.261)	(0.235)	(0.260)	(0.240)
ln(D_PM2.5)	-0.369	-0.978***	-0.530**	-1.193***	0.100	-0.417**
	(0.246)	(0.207)	(0.249)	(0.210)	(0.253)	(0.212)
ln(O_BOD)	-	-0.0469	-	-0.0862	-	0.0308
	-	(0.101)	-	(0.102)	-	(0.102)
ln(D_BOD)	-	0.904***	-	0.940***	-	0.798***
	-	(0.124)	-	(0.123)	-	(0.124)
ln(O_GWL)	-0.0573	-0.463***	-0.135	-0.431***	0.0445	-0.528***
	(0.353)	(0.102)	(0.349)	(0.101)	(0.345)	(0.106)
ln(D_GWL)	3.264***	-0.195	3.310***	-0.293**	3.092***	-0.174
	(0.576)	(0.148)	(0.580)	(0.147)	(0.618)	(0.151)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Yes	Yes	Yes	Yes	Yes	Yes
F statistic	28.729	43.230	28.780	42.994	24.342	42.411
N	6,121	6,121	5,898	5,898	5,731	5,731
R^2	0.200	0.486	0.110	0.433	0.189	0.474

Notes: The standard errors in parenthesis are clustered at origin state-destination district pair level. Controls include typical gravity variables of populations and distances, social variables and weather variables. "O_" and "D_" stand for origin and destination, respectively. ***, ** and * indicates statistical significance at 1%, 5% and 10% level respectively.

6.2 Robustness Check and Heterogeneity

A series of alternative regressions were run to check robustness of our results. First, standard errors are clustered at state-pair level to account for potential within-state-pair autocorrelation across states and over time periods (Table 18-19 in Appendix). In our baseline, standard errors are clustered at state-district level. Unfortunately, standard errors are much larger than baselines and some coefficients lose their statistical significance. Nevertheless, destination groundwater coefficients for all gender categories are still significant at 5% level. Second, Table 20-21 show the results using one-year values of environmental factors instead of three-years average. As discussed previously, the full effects of environmental factors on migration may be delayed over time, thus baseline captures mid-term delayed effects. Instead, we here examine the immediate effect on migration. The estimates are robust to the alternative main variables and suggest the consistency of our findings.

Third, due to time-invariant nature of GWL instrument, the baseline uses only period FE to control for common shocks across the country. This may cause a failure to control for spatial time-fixed differences of environmental factors within origin and destination. However, gravity model structure of our economic framework is not compatible with the inclusion of both origin and destination FE since it ends up with omitting all of our main explanatory variables GWL. Hence, we include either origin or destination FE at once. By doing so, we can constrain the instrument-related variation in each environmental factor to deviation from the within-state or within-district average value of the instrument. Also, more stringent origin-period FE and destination-period FE are added for all environmental factors, and demanding state-district pair FE is attempted for air and water pollution. Origin and destination FE can control factors including geographic-specific characteristics such as labour market condition, educational opportunity, elevation and aridity of land as well as relationships with the third areas which is so-called multilateral resilience in gravity model. Pair-wise FE can control for factors such as historical relationships. Table 22-25 in Appendix suggests that the results are robust to additional state FE and/or state-time FE or even pair-wise FE. When we control for destination FE, the coefficients tend to lose its statistical and economic significance. On one hand, as more FE controls are added, the fitness of the model clearly improves, indicating that each FE has some extent of values in terms of explanatory powers of the equation. However, there is an inherent threat of overfitting from adding high dimensional fixed effects. If overfitting happens, then some variables start to explain error term, resulting in less predictive power of the model with misleading coefficients, R-squared and p-values.

Table 10: Percentage of Migrants by Area Types

	Both genders		Male		Female	
	2001	2011	2001	2011	2001	2011
Rural-Rural	0.458	0.436	0.448	0.365	0.471	0.492
Rural-Urban	0.249	0.216	0.260	0.248	0.235	0.189
Urban-Rural	0.093	0.119	0.107	0.136	0.073	0.106
Urban-Urban	0.174	0.228	0.160	0.249	0.193	0.211
Total	0.974	0.999	0.975	0.998	0.972	0.998

Notes: The Table 0 presents the breakdown of total migrants by area types of origin and destination. Since the survey did not distinguish the origin types for a handful group of people due to no indication by surveyees, the sums of 4 categories are slightly less than 1.

Fourth, another aspect is potential heterogeneity across migration types. The data distinguish rural or urban areas of their residence at origin and destination. Due to unbalanced counts of each combination shown in Table 10, we split the dataset into two groups: the one who migrated from rural area and the other who migrated from urban area. The distinction of origin area types can be a rough indicator for the income level or job types of migrants, capturing demographic heterogeneity. Table 26 in Appendix presents that migrants from urban areas are interestingly much more sensitive to air pollution. Intuitively, it makes sense as individuals living urban areas are likely to be wealthy enough to pay more attention to health impact of inhaling pollutants. Greenstone and Jack (2015) indeed finds that poorer people have lower value for the willingness to pay for environmental quality, due to lower priority than income or today's livelihood. In contrast, the economic magnitude of GWL for migrants from rural area is more than double of that of migrants from urban areas. This is also intuitive and consistent with the findings of Zeveri et al. (2020) saying that individuals in rural areas are sensitive to water access and likely to send migrants from family because of the reason.

7 Conclusion

This paper investigates the impact of air pollution, water pollution and water scarcity on internal migration in India using gravity model with instrumental variable approach. To the best our knowledge, previous literature has not well covered these areas, especially in Indian context. This study is the first to examine the relationship by incorporating nationwide migrants and three environmental factors. The dataset is constructed from a wide range of sources including Indian government data platforms and satellite data. We use thermal inversion, rainfall and rock types as instruments for each environmental indicator. Our analysis finds that a positive and statistically significant effect of air pollution at origin working as a push factor. The estimates favourably provide a confirmatory evidence to previous research in terms of the direction of impacts. On the contrary, our results imply that decrease in GWL as a proxy for water scarcity induces migration at origin while the increase in GWL at destination pulls in-migrants across gender. In particular, the latter impact of water scarcity is robust to various types of econometric specifications.

However, we do not find consistent and significant evidence on water pollutants. Potential reasons are not comprehensive coverage of data and endogeneity bias stemming from failure to control weather and economic variables, raising doubt on our IV. Another concern is that our study period is rather short and not ideal provided that environmental issues internationally became more urgent and the importance of health risks are more aware of in the last decade. Also, our environmental proxies measure each issue from single scope, in turn, there is likely to exist measurement errors inherent in the parameters. Moreover, our data do not allow us to distinguish total migrants by their educational background nor job categories. Heterogeneity analysis from such aspect gives us further policy implications. The validity of our results is limited in those regards. Nevertheless, our findings provide environmental policy implications especially in developing countries, that pollutions and water scarcity are significant burdens for residents inducing their geographical movement. Such spatial sorting would lead to inefficient distribution of human resources in labour markets. In particular, water scarcity is an immediate issue especially for individuals relying on agriculture not only affecting migration flows but also inducing significant repercussions on overall economy. With the aim of deeper understanding, future studies can investigate how the effect of pollutions and water scarcity varies across job types of household.

8 Bibliography

Anderson, J. and van Wincoop, E., 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1), pp.170-192.

Anderson, S., 2011. Caste as an Impediment to Trade. *American Economic Journal: Applied Economics*, 3(1), pp.239-263.

Angrist, J. and Pischke, J., 2009. *Mostly Harmless Econometrics*. Princeton University Press.

Arceo, E., Hanna, R. and Oliva, P., 2016. Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City. *The Economic Journal*, 126(591), pp.257-280.

Beine, M. and Parsons, C., 2014. Climatic Factors as Determinants of International Migration. *The Scandinavian Journal of Economics*, 117(2), pp.723-767.

Bhattacharya, P., 2002. Rural-to-urban migration in LDCS: a test of two rival models. *Journal of International Development*, 14(7), pp.951-972.

Borjas, G.J., 1999. The Economic Analysis of Immigration, *Handbook of Labor Economics*, in: Ashenfelter, O. and Card, D. (eds.). pp.1697-1760.

Burgess, R., Deschenes, O., Donaldson, D. and Greenstone, M., 2017. Weather, Climate Change and Death in India. Working paper, London School of Economics.

Central Ground Water Board (CGWB). 2019. Ground Water Year Book of States 2018-19. Department of Water Resources, River Development and Ganga Rejuvenation, Ministry of Jal Shakti, Government of India, Faridabad

Chandramouli, C., 2013. *Release of Primary Census Abstract Data Highlights*. [online] Available at: <https://idsn.org/wp-content/uploads/user_folder/pdf/New_files/India/2013/INDIA_CENSUS_ABSTRACT-2011-Data_on_SC-STs.pdf> [Accessed 22 July 2021].

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

Chay, K. and Greenstone, M., 2004. Does Air Quality Matter? Evidence from the Housing Market. *SSRN Electronic Journal*,.

Chen, S., Oliva, P. and Zhang, P., 2017. The Effect of Air Pollution on Migration: Evidence from China. *SSRN Electronic Journal*,.

Clark, X., Hatton, T., and Williamson, J., 2007. Explaining U.S. Immigration, 1971-1998. *The Review of Economics and Statistics*, 89(2), pp.359-373.

Central Pollution Control Board. 2013. *Status of Water Quality in India*. [online] Available at: <http://nwm.gov.in/sites/default/files/waterwiki/1.pdf> [Accessed 9 August 2021].

Central Pollution Control Board. 2021. *CPCB ENVIS| Control of Pollution*. [online] Available at: <http://cpcbenvis.nic.in/water_quality_data.html> [Accessed 4 August 2021].

Cui, C., Wang, Z., He, P., Yuan, S., Niu, B., Kang, P. and Kang, C., 2019. Escaping from pollution: the effect of air quality on inter-city population mobility in China. *Environmental Research Letters*, 14(12), pp.124025.

Dallmann, I. and Millock, K., 2017. Climate Variability and Inter-State Migration in India. *CESifo Economic Studies*, 63(4), pp.560-594.

Dell, M., Jones, B.F. and Olken, B.A., 2014. What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, 52(3), pp.740-798.

Dubey, A., Palmer-Jones, R. and Sen, K., 2006. Surplus labour, social structure and rural to urban migration: Evidence from Indian data. *The European Journal of Development Research*, 18(1), pp. 86-104.

Feenstra, R., 2002. Border Effects and the Gravity Equation: Consistent Methods for Estimation. *Scottish Journal of Political Economy*, 49(5), pp.491-506.

Fishman, R., Jain, M. and Kishore, A., 2013. Patterns of Migration, Water Scarcity and Caste in Rural Northern Gujarat. [online] Theigc.org. Available at: <https://www.theigc.org/wp-content/uploads/2015/02/Fishman-Et-Al-2013-Working-Paper.pdf> [Accessed 22 June 2021].

Flowerdew, R. and Aitkin, M., 1982. A Method of Fitting the Gravity Model Based on the Poisson Distribution. *Journal of Regional Science*, 22(2), pp.191-202.

Freeman, R., Liang, W., Song, R. and Timmins, C., 2019. Willingness to pay for clean air in China. *Journal of Environmental Economics and Management*, 94, pp.188-216.

Geological Survey of India. 2021. <lithology>, Geological Survey of India. Government of India, Kolkata, India. 10 August 2021.

Germani, A., Scaramozzino, P., Castaldo, A. and Talamo, G., 2021. Does air pollution influence internal migration? An empirical investigation on Italian provinces. *Environmental Science & Policy*, 120, pp.11-20.

Global Modeling and Assimilation Office (GMAO). 2015. MERRA-2 inst6_3d_ana_Np: 3d,6-Hourly, Instantaneous, Pressure-Level,Analysis,Analyzed Meteorological Fields V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [*17/08/2021*], 10.5067/A7S6XP56VZWS

Government of India., 2011. Provisional Population Totals. Census of India 2011, Paper 1 of 2011, pp.188.

Government of India. 2017a. India on the Move and Churning. In Economic Survey of India 2016–17, 264–84. New Delhi: Ministry of Finance.

Hicks, D., Marsh, P., and Oliva, P., 2015. Air pollution and procyclical mortality: Causal evidence from thermal inversions. Paper presented at the paper session of Pollution and Later-Life Mortality hosted by American Economic Association, 6 January 2017, Chicago, Illinois, USA.

Imbert, C. and Papp, J., 2020. Short-term Migration, Rural Public Works, and Urban Labor Markets: Evidence from India. *Journal of the European Economic Association*, 18(2), pp.927-963.

Khanna, G., Liang, W., Mobarak, A. and Song, R., 2021. The productivity consequences of pollution-induced migration in China | VOX, CEPR Policy Portal. [online] Voxeu.org. Available at: <https://voxeu.org/article/productivity-consequences-pollution-induced-migration-china> [Accessed 21 June 2021].

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

Li, D., Zhang, Y. and Ma, S., 2017. Would Smog Lead to Outflow of Labor Force? Empirical Evidence from China. *Emerging Markets Finance and Trade*, 53(5), pp.1122-1134.

Li, X., Chen, H. and Li, Y., 2020. The effect of air pollution on children's migration with parents: evidence from China. *Environmental Science and Pollution Research*, 27(11), pp.12499-12513.

Liu, Z. and Yu, L., 2020. Stay or Leave? The Role of Air Pollution in Urban Migration Choices. *Ecological Economics*, 177, pp.106780.

Mannucci, P. and Franchini, M., 2017. Health Effects of Ambient Air Pollution in Developing Countries. *International Journal of Environmental Research and Public Health*, 14(9), pp.1048.

Marchiori, L., Maystadt, J. and Schumacher, I., 2012. The impact of weather anomalies on migration in sub-Saharan Africa. *Journal of Environmental Economics and Management*, 63(3), pp.355-374.

McKee, T., Doesken, N. and Kleist, J., 1993. The relationship of drought frequency and duration to time scales. Paper presented at the 8th Conference on Applied Climatology, January 17-22, Anaheim, California, USA.

McLeman, R. and Smit, B., 2006. Migration as an Adaptation to Climate Change. *Climatic Change*, 76(1-2), pp.31-53.

Mitra, A. and Ray, D., 2014. Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India. *Journal of Political Economy*, 122(4), pp.719-765.

Muñoz Sabater, J., 2019. ERA5-Land Monthly Averaged Data from 1981 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). [Accessed 21 June 2021]. 10.24381/cds.68d2bb3.

Munshi, K. and Rosenzweig, M., 2016. Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review*, 106(1), pp.46-98.

Neidell, M., 2004. Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of Health Economics*, 23(6), pp.1209-1236.

Panda, D., Mishra, A. and Kumar, A., 2012. Quantification of trends in groundwater levels of Gujarat in western India. *Hydrological Sciences Journal*, 57(7), pp.1325-1336.

Pant, P., Guttikunda, S. and Peltier, R., 2016. Exposure to particulate matter in India: A synthesis of findings and future directions. *Environmental Research*, 147, pp.480-496.

Pischke, S., Lecture Notes on Measurement Error. Working paper, London School of Economics.

Prabha, A., Ram, A. and Irfan, Z., 2020. Exploring the relative water scarcity across the Indian million-plus urban agglomerations: An application of the Water Poverty Index. *HydroResearch*, 3, pp.134-145.

Rajan, S. and Neetha, N., 2018. *Migration, Gender and Care Economy*. Oxen and New York: Routledge India.

Sekhri, S. and Hussain, M., 2019. Water in Scarcity, Women in Peril. [online] Blogs.darden.virginia.edu. Available at:

<https://blogs.darden.virginia.edu/globalwater/2019/09/09/water-in-scarcity-women-in-peril/> [Accessed 21 June 2021].

Seiler, R., Hayes, M. and Bressan, L., 2002. Using the standardized precipitation index for flood risk monitoring. *International Journal of Climatology*, 22(11), pp.1365-1376.

Sekhri, S., 2014. Wells, Water, and Welfare: The Impact of Access to Groundwater on Rural Poverty and Conflict. *American Economic Journal: Applied Economics*, 6(3), pp.76-102.

Shannon, A., Usmani, F., Pattanayak, S. and Jeuland, M., 2018. The Price of Purity: Willingness to Pay for Air and Water Purification Technologies in Rajasthan, India. *Environmental and Resource Economics*, 73(4), pp.1073-1100.

Silva, J. and Tenreyro, S., 2006. The Log of Gravity. *Review of Economics and Statistics*, 88(4), pp.641-658.

Xue, S., Zhang, B. and Zhao, X., 2019. Brain Drain: The Impact of Air Pollution on Firm Performance. *SSRN Electronic Journal*.

Wada, Y., van Beek, L., van Kempen, C., Reckman, J., Vasak, S. and Bierkens, M., 2010. Global depletion of groundwater resources. *Geophysical Research Letters*, 37(20).

Warziniack, T., 2013. The Effects of Water Scarcity and Natural Resources on Refugee Migration. *Society & Natural Resources*, 26(9), pp.1037-1049.

Wilson, A., 1971. A Family of Spatial Interaction Models, and Associated Developments. *Environment and Planning A: Economy and Space*, 3(1), pp.1-32.

World Bank. 2010. Deep wells and prudence: Towards pragmatic action for addressing groundwater overexploitation in India. World Bank Technical Report, The World Bank.

World Bank. 2016. *High and Dry: Climate Change, Water, and the Economy*. [online] Available at: <https://www.worldbank.org/en/topic/water/publication/high-and-dry-climate-change-water-and-the-economy> [Accessed 3 July 2021].

World Bank. 2019. *Helping India Overcome Its Water Woes*. [online] Available at: <https://www.worldbank.org/en/news/feature/2019/12/09/solving-water-management-crisis-india> [Accessed 14 August 2021].

World Health Organization. 2014. *7 Million Premature Deaths Annually Linked to Air Pollution*. [online] Available at: <https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution> [Accessed 28 June 2021].

van Donkelaar, A., R. V. Martin, M. Brauer, N. C. Hsu, R. A. Kahn, R. C. Levy, A. Lyapustin, A. M. Sayer, and D. M. Winker., 2018. Global Annual PM2.5 Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD) with GWR, 1998-2016. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Available at: < https://doi.org/10.7927/H4ZK5DQS > [Accessed 21 June 2021].