# Climate Variability and Internal Migration: A Test on Indian Inter-State Migration

Ingrid Dallmann, Katrin Millock

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

# Climate Variability and Internal Migration: A Test on Indian Inter-State Migration

Ingrid DALLMANN, Katrin MILLOCK

**2013.45R**
*Version révisée*

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/
ISSN : 1955-611X

# Climate Variability and Internal Migration: A Test on Indian Inter-State Migration *

Ingrid Dallmann† Katrin Millock‡

March 25, 2016

**Abstract**

We match migration data from the 1991 and 2001 Indian Census with climate data to test the hypothesis of climate variability as a push factor for internal migration. The article contributes to the literature by combining three important factors. First, we introduce relevant meteorological indicators of climate variability, based on the standardized precipitation index. Second, the use of the census data enables us to match the migration data with the relevant climate data ex ante, rather than relying on average conditions. Third, we analyse bilateral migration rates in order to fully account for characteristics in both the origin and the destination. We therefore use an econometric estimation method that accounts for zero observations, which are frequent in bilateral data. The estimation results show that drought frequency in the origin state acts as push factor on inter-state migration in India. We do not find a statistically significant effect of the magnitude and the duration of drought episodes preceding migration. There is no evidence of excess precipitation acting as a push factor on inter-state migration. The results are robust to alternative specifications of fixed effects and to the inclusion of irrigation rates.

*JEL codes:* O15, Q54.

*Keywords:* climate variability, drought, India, internal migration, PPML, SPI.

*We thank Catherine Bros, Sudeshna Chattopadhyay, Miren Lafourcade and Brinda Viswanathan for their help and advice, as well as participants in seminars at the universities of Gothenburg, Helsinki, Orléans and the 2nd International Conference on Environment and Natural Resources Management in Developing and Transition Economies (enrmdte), in particular Simone Bertoli. Any errors or omissions are only the authors' responsibility, naturally. Financial support from the French National Research Agency grant ANR-JCJC-0127-01 is gratefully acknowledged.

†University Paris-Sud 11 (RITM); ingrid.dallmann-gamarra@u-psud.fr

‡Paris School of Economics, CNRS, Centre d'Economie de la Sorbonne; millock@univ-paris1.fr

# 1 Introduction

Negative effects linked to climate variability are more and more apparent, not only through the increase in natural disasters that cause huge economic and human losses but also through its long-term consequences on the economy and on the population distribution. The most recent assessment report by the Intergovernmental Panel on Climate Change (IPCC, 2014) shows ample evidence on the different manners in which climate change will affect migration, although it is difficult to quantify the expected flows. Detailed studies in a report commissioned by the UK government (Government Office for Science, 2011) show that environmental change will affect migration in the present and in the future, but that the influence will be principally through economic, social and political drivers. Climate variability, in particular, may have direct effects, such as degraded health, death, capital destruction and disruption of socio-economic activities, but also have indirect effects on the environment and the economy, through price and wage adjustments in the market, hence inducing migration either directly or indirectly. The purpose of this paper is to test the hypothesis that climate variability acts as a push-factor on internal migration.

The article makes three major contributions to the existing literature on macro-level migration flows. First, the advantage of the current study is to use the one-year migration definition from the Indian Census that permits an exact match between climatic factors prior to migration and the observed migration flow.This causality is not present in existing studies that rely on average migration flows over a 5- or 10-year period linked to average climate anomalies over the same 5-10 year period. Second, we introduce relevant meteorological indicators of climate variability based on the Standardized Precipitation Index (SPI). The SPI measures anomalies in rainfall compared to a long-run average defined from 1901 up to the year of the census. It is normalized and hence comparable across different locations with different climate. Our climate factors measure the frequency, magnitude and duration of droughts in one- to five- year periods preceding the migration. Similar measures are constructed for excess precipitation. We use the term climate variability rather than weather since we measure deviations from the long run mean based on the standardized distribution over a century of observations. Third, we use bilateral migration rates in order to control for important existing migration determinants between states. This necessitates the use of an econometric estimation method that accounts for zero observations, which are frequent in bilateral data. We also account for the endogeneity of income to precipitation and exclude the income variable from the estimations, contrary to previous work. We hence estimate the net effect on migration, without separating the direct and the indirect effects of climate variability.

We match bilateral migration data from the 1991 and 2001 Indian Census with climate data at the state level from the Climatic Research Unit at the University of East Anglia (CRU). We use the SPI to measure variability. The estimation results show that the frequency of drought has a significant impact on inter-state bilateral migration rates once we control for other natural disasters, criminality, migration costs, networks and all the destination-state pull factors. Each additional month of drought in the origin state during the five years preceding the year of migration increases the bilateral migration rate by 1.6%. The relative effect of climate variability is rather small, though, compared to the effect from migration costs as measured by barriers to inter-state migration. We find no evidence of excess precipitation acting as a push factor on inter-state migration. The results are robust to controlling for irrigated land in the state and also to the inclusion of controls for all bilateral fixed effects invariant in time.

We contribute to a growing literature that analyses the link between migration and climate variability. The idea that negative environmental conditions would increase international migration was popularized in the "environmental refugees" literature (Myers, 1997), but since reinterpreted and moderated by Piguet (2010) and Gemenne (2011), amongst others. On the one hand, several studies use detailed micro economic data to analyse factors linking migration with climatic conditions. For example, in a large household study on Bangladesh, Gray and Mueller (2012) find that floods did not have a significant impact on migration, whereas weather-related crop failure did. In another study relating counts of natural disasters with migration inferred from the Indonesian Family Life Surveys, Bohra-Mishra et al. (2014) find no significant impact from natural disasters other than landslides on internal migration of the entire household, but a significant and large effect of temperature and a significant but smaller effect of rainfall. The advantage of this literature, reviewed in Lilleor and Van den Broeck (2011), is to show how individual household factors contribute to vulnerability and to explain what makes some households migrate whereas others will not. Nevertheless, it is difficult to generalize the findings of these studies outside of the particular country analysed.

On the other hand, macroeconomic studies on international migration flows such as Reuveny and Moore (2009), Beine and Parsons (2015) and Coniglio and Pesce (2015) aim at testing the effect on cross-border flows. Reuveny and Moore (2009) show that both weather-related natural disasters and climate anomalies may (directly) induce increased migration into OECD countries. In a comprehensive study of international migration over the period 1960-2000, Beine and Parsons (2015) find no effect of either temperature or rainfall deviations on international bilateral migration flows, including south-south migration. Coniglio and Pesce (2015) test additional definitions of weather variables and find evidence of a positive effect of the inter-annual variability of rainfall on out-migration to OECD countries. This difference in conclusions stems partly from the use of different datasets: migration flows calculated from migration stock data at 10 year intervals from 1960 to 2000 in Beine and Parsons (2015) and annual data over a shorter time span (1990-2001) in Coniglio and Pesce (2015). When long-term migration average data is used, it is difficult to match exactly climatic factors with the observed migration flows in order to establish causality.

The current article is part of a group of recent analyses on climatic factors and migration relying on the most comprehensive data for migration flows at a country level, i.e., census data. Few studies use census data to study climatic factors and internal migration in large countries, and they are mainly from the U.S. (Boustan et al., 2012; Feng et al., 2012). Whereas Feng et al. (2012) study the indirect effect of temperature-induced crop shocks on out-migration from the U.S. corn belt states, Boustan et al. (2012) show that floods and tornados had a significant effect on gross migration flows in the U.S. in the 1920s and 1930s.

We focus on internal (inter-state) migration in India, since migration induced by climate variability is more likely to occur within the internal borders of a country, because of migration costs, including legal barriers (Marchiori et al., 2012; Beine and Parsons, 2015). In addition, low-income and lower-middle-income countries are also more vulnerable to climate variability than high-income countries (Stern, 2007; Government Office for Science, 2011) due to their lower adaptation capacity and their geographical location. In order to fully account for all possible factors influencing migration, bilateral flows should be used, and unfortunately, this prohibits us from using more detailed data on a district level, since the origin of migrants is not recorded at this level, but only the destination. We undertake the first comprehensive study of internal bilateral migration rates and climate variability on a country as large and diverse as India and introduce new standardized exogenous measures of climate variability that allow for coherent comparisons across states with very different climatic conditions. In doing so we also contribute to the migration literature that typically uses gravity-type models that only incorporate socio-economic factors but not environmental ones (Karemera et al., 2000, Mayda, 2010, Van Lottum and Marks, 2010, and in particular Özden and Sewadeh, 2010 on India).

The only other studies on migration and climate in India either analyse cross-section data from the National Sample Survey (NSS), as in Kumar and Viswanathan (2013), or use census data to apply the method of Feng et al. (2012) to study migration induced by agricultural shocks only (Viswanathan and Kumar, 2015). By contrast, we aim at measuring the total effect on internal migration, encompassing both direct effects on utility, such as health impacts, and indirect transmission through income effects. Another difference is the use of the complete census data (31 out of the 32 states according to the 1991 state borders) for 1991 and 2001, whereas Viswanathan and Kumar (2015) analyse data from the 15 major states, but over the period 1981-2001. Their state-level analysis shows that weather-induced shocks to agricultural income induce out-migration for employment purposes. We control for all possible factors that can affect bilateral migration flows, which is important to identify an effect of climatic factors on out-migration.

The remainder of the paper proceeds as follows. Section 2 presents some statistics about climate variability and inter-state migration in India. Section 3 and 4 discuss the empirical strategy of estimation and the data employed respectively. Section 5 shows the empirical results, and Section 6 concludes.

## 2 Inter-state migration and climate variability in India

Analyzing inter-state migration in India is particularly appropriate for a study of internal migration because of the heterogeneity among states, especially as regards demography and climate. India has a large variety of climate regions, ranging from tropical in the South to temperate and alpine in the Himalayan North. The main natural disasters in India are drought, flood and tropical cyclones, in order of the number of people affected (Attri and Tyagi, 2010). In this analysis, we focus on droughts and floods. India is also considered by the Environmental Vulnerability Index[1] as extremely vulnerable, not only because of its climate vulnerability, but also because of

[1]Index developed by the South Pacific Applied Geoscience Commission (SOPAC) and the United Nations Environment Program (UNEP).

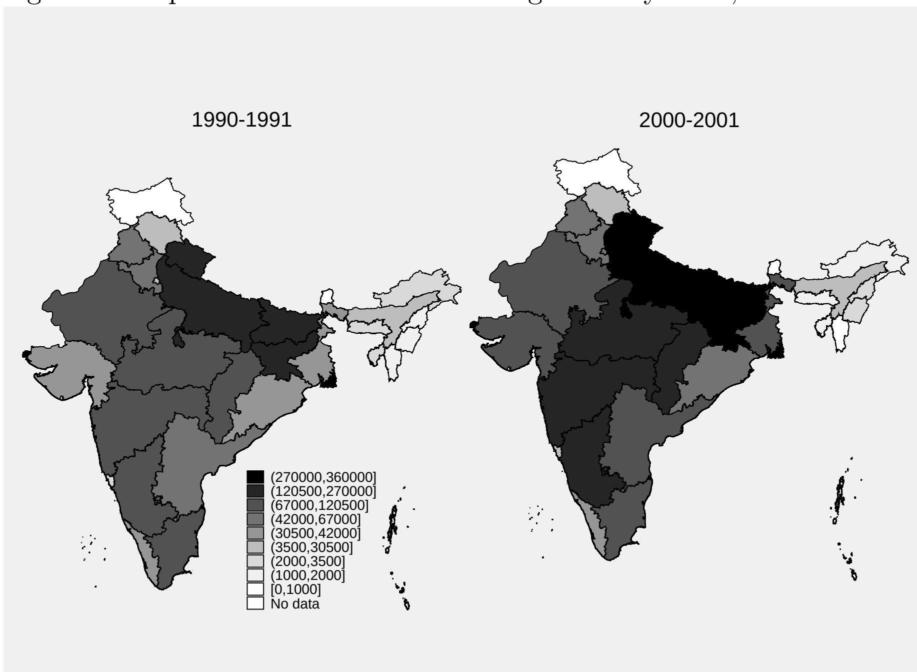
its population density. India is indeed the second most populated country in the world with 1,210 million inhabitants in 2011 that represents 17.5% of the world population with only 2.4% of the world surface area, and a population growth between 2001 and 2011 of 17.6%. Its population is mainly rural, of 69% in 2011, representing 833.5 million people (Census of India, 2011).Population densities differ much between states, for instance it ranges from 17 to 11,297 persons per square km in 2011 (Arunachal Pradesh and Delhi respectively). In 1991, 26.7 % of the total population was an internal migrant, of which 11.8 % were inter-state migrants. In 2001, this proportion increased to 30.1% (310 million persons) with 13.4% of the migrants being inter-state migrants. International migration is only 3.8 % in India, according to the 64th round of the National Sample Survey (NSS) conducted in 2007-2008 (Czaika, 2011). These statistics motivate the interest in analysing the potential influence of climate variability on internal migration.

We use the definition of migrants as individuals declaring the last place of residence in year $t-1$ to be different from the place of enumeration in the years 1991 and 2001. The use of the last year's migration flow data enables us to match the data more precisely in time with the climatic factors. This is an advantage compared to the existing literature on climatic factors and migration, that often is based on less precise estimations calculated on 5 or 10 year averages of migrant stock data.

Figure 1 shows the number of out-migrants by state in 1990-1991 and 2000-2001. It confirms the description in Özden and Sewadeh (2010) of the major North-Western migration corridors based on data from the 55th round of the NSS in 1999-2000. The states with the highest numbers of inter-state out-migrants are the Northern states Uttar Pradesh and Bihar, the Central state Madhya Pradesh and the South-Western states Maharashtra and Karnataka (in dark colours).

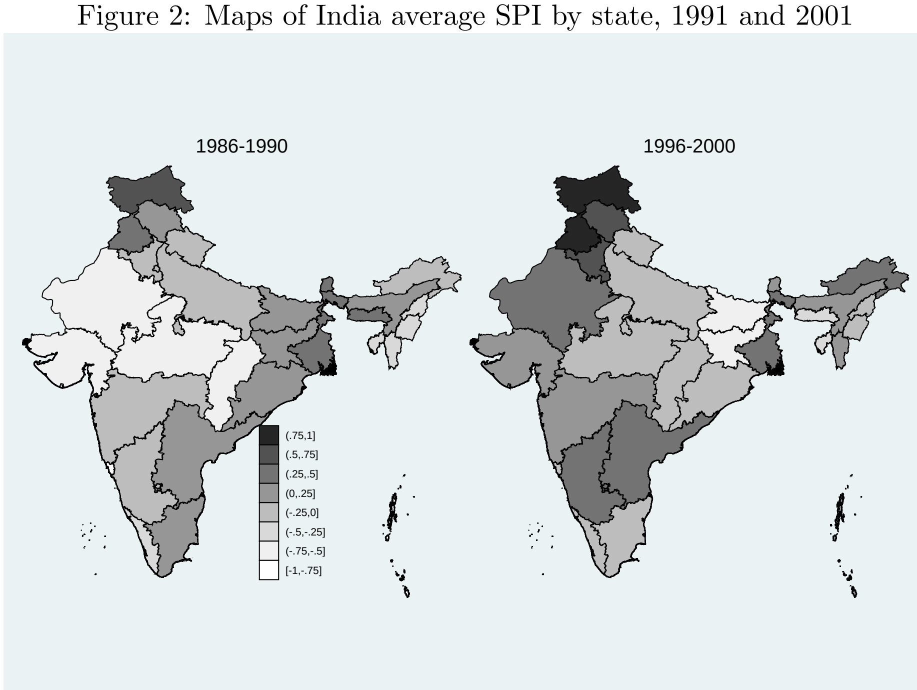
Figure 2 shows the average SPI for the five years preceding the migration flow (1986-1990 and 1996-2000) for illustrative purposes. It ranges from -1 to +1, which represents moderate deviations, because of smoothing over time. The lighter colours indicate negative values, and thus a deficit of precipitation compared to the long run mean, whereas the darker colours indicate excess precipitation. A comparison of the two maps shows that the major out-migration states all had negative values of the SPI, on average, before 1991. For the South-Western states Karnataka and Maharashtra the average SPI returned to around zero in 2001, though, whereas for Bihar the average SPI became more negative.

Figure 1: Maps of India interstate out-migration by state, 1991 and 2001




*Source:* Indian Census 1991 and 2001, D2-Series.
The definition of migrants is that of individuals declaring the last place of residence in year $t-1$ to be different from the place of enumeration in year $t$ in the Census.

Figure 2: Maps of India average SPI by state, 1991 and 2001




*Source:* Authors' own calculations based on the CRU TS3.21 data.

# 3 Empirical specification and method

## 3.1 Theoretical framework and econometric specification

We use a standard migration model in economics (as in Beine et al., 2011, or Beine and Parsons, 2015) to specify the determinants of bilateral migration and control for both "push" and "pull" factors of migration:

$$\ln \frac{m_{ij,t}}{m_{ii,t}} = \ln \frac{w_{j,t}}{w_{i,t}} + S_{j,t} - S_{i,t} - C_{ij,t} \tag{1}$$

where $m_{ij,t}$ is the bilateral migration rate from state $i$ to state $j$, and $m_{ii,t}$ is the stock of population in state $i$ at time $t$. The determinants are the income in the state of origin ($w_{i,t}$), the income in the destination state ($w_{j,t}$ ), the origin state characteristics ($S_{i,t}$), time-varying destination state characteristics ($S_{j,t}$), and the cost of migration between the two states at time $t$ ($C_{ij,t}$ ).

Since income is endogenous to climate (Dell et al., 2009; Burke et al., 2015) we cannot include it. As argued in Dell et al. (2014), there is a trade-off between omitted variable bias and over controlling by including endogenous regressors in the estimation. Indeed, we find that the income ratio is significantly affected by the climatic factors (see Table C.5 in Appendix C), and exclude the income ratio from the estimation since it would bias the measure of the net effect of climatic variables on migration. Instead we estimate equation 1 directly including only climatic factors and other state time varying characteristics together with state fixed effects. The resulting specification (in equation 2) thus captures the net effect of climate variability on bilateral migration rates.

The cost of migration is represented by distance between state $i$ and state $j$ ($d_{ij}$), and dummy variables for common border ($b_{ij}$) and language ($l_{ij}$) between states, as is common in migration analyses (Bodvarsson and Van den Berg, 2009).[2] We also control for caste (or ethnic) similarity between states by including the ratios of scheduled castes and scheduled tribes in the destination state compared to the origin state. In India, 16.2 % of the population belong to a scheduled caste (SC), also called "the untouchables", and 8.2% to scheduled tribes (ST) in 2001. In the literature on Indian migration, these two factors are almost always taken into account to examine the role of social factors in the migration decision (Bhattacharya, 2002, and Mitra and Murayama, 2008). The "Hindu Varna" System, which establishes the classification of the society in India, categorizes groups of population on the basis of the caste, the ethnicity and the religion. This discrimination persists in the labour force participation (Dubey et al., 2006). SC and ST may be the most vulnerable parts of the population to climate variability given that they often are day labourers and hence likely to be the first affected by climate events. In particular, they have less access to water resources that may be drawn upon in times of drought (groundwater sources, water tank sales), as shown by Anderson (2011). If these groups of individuals experience discrimination

[2]See Appendix A for a detailed description of all data.

from upper castes and dominant groups, we may hypothesize that they would like to stay within their communities and be more likely to migrate (if they do so) where they can find their peers. Indeed, Bhattacharya (2002) find that scheduled castes are less likely to migrate (from rural to urban areas) but if they do so, they go where they can find other scheduled caste population. We include the natural logarithm of the ratios of the scheduled caste ($\frac{SC_{j,t}}{SC_{i,t}}$) and tribe rates ($\frac{ST_{j,t}}{ST_{i,t}}$) in the destination state compared to the origin state to control for this.

Time-varying origin state characteristics, $S_{i,t}$, that may affect migration rates are other natural disasters, violence, and possibilities to adapt to climate variability. We would like to control for exogenous natural disasters other than climate-driven ones (such as cyclones and other natural disasters), but the data we studied from the Emergency Events Database (EM-DAT), collected by the Centre for Research on the Epidemiology of Disasters (CRED) at the Université Catholique de Louvain, did not seem reliable at the state level, as compared to the country level. We include an indicator variable that takes the value of one if there were any earthquake recorded in the origin state in the year prior to migration ($earthquake_{i,t-1}$). Violence could induce migration, in particular religious or caste-based violence (Mitra and Raj, 2014). Lacking data on these particular origins of violence for all 31 states in the census, we control for total murder rates by 100,000 inhabitants in the state of origin one year before migration ($murder_{i,t-1}$).

The principal variables of interest are the ones representing variability in precipitation and the duration and magnitude of a period with low or excess precipitation ($clim_{i,t}$). Our hypothesis is that variability in precipitation and adverse weather events act as a push factor on migration. In particular, this is the case in developing countries where poor people do not move by comparing origin and destination climatic factors but rather escape from drought or floods that affect their well-being. Accordingly, all our variables representing variability and adverse weather events act only in the origin state. We count deviations in precipitation over a five-year period preceding migration, since the full effect of such events on migration may be delayed over time.

We include origin state fixed effects ($D_i$) that are invariable in time to capture the vulnerability of the geographic zone, especially mountains, low elevation coasts and arid lands. This dummy controls also for the states affected by the Armed Forces (Special Powers) Act of 1958. The Act gives special power to armed forces (military and air forces) in the so called "disturbed" areas. The states and Union Territories affected are: Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland and Tripura. These states have experienced violence that may have induced migration. Migration also varies according to employment perspectives on the labour market in the destination, and can also be due to educational opportunities. All such time-varying characteristics of the destination state are captured through destination state and time fixed effects ($D_{j,t}$) including potential climate pull effect.

The resulting econometric specification thus estimates the migration flow from state $i$ to state $j$ during year $t-1$ to $t$ as a rate of the initial population staying in state $i$ during year $t-1$ to $t$:

$$
\begin{aligned}
\ln \frac{m_{ij,t}}{pop_{ii,t}} = & a_0 + a_1 \ln d_{ij} + a_2 b_{ij} + a_3 l_{ij} + \\
& + a_4 \ln \frac{SC_{j,t}+1}{SC_{i,t}+1} + a_5 \ln \frac{ST_{j,t}+1}{ST_{i,t}+1} \\
& + a_6 \ln murder_{i,t-1} + a_7 earthquake_{i,t-1} \\
& + a_8 \sum_{t-5}^{t} clim_{i,t} + D_i + D_{j,t} + u_{ij,t}
\end{aligned}
\tag{2}
$$

The expected signs on the variables representing the costs of migration are: $a_1 < 0$, $a_2 > 0$, $a_3 > 0$. The relation of migration with distance is negative, as a proxy of migration travel costs. Common border and language reduce the cost of migration, since they are proxies of cultural similarities between states.

The SC and ST variables constitute a wider measure of migration costs in the sense that, for an individual belonging to the SC or ST population, moving to a state with a higher ratio of SC (or ST) compared to the origin state would imply lower costs of migration because of the network in the destination state, whereas moving to a state with a lower ratio of SC (or ST) would imply higher costs of migration because of the smaller network. Ex ante, the coefficients $a_4$ and $a_5$ could thus be either positive or negative.

We assume that higher murder rates, all else equal, will induce more migration and expect $a_6 > 0$. For earthquakes, the impact is ambiguous ex ante. One may expect that natural disasters induce migration and that $a_7 > 0$, but Hallyday (2006, 20012) finds that the 2001 earthquakes in El Salvador had no significant effect on male out-migration, and significantly decreased female out-migration.

For the variables representing variability in precipitation we expect a positive sign ($a_8 > 0$) for the measures of lower than average precipitation since the SPI is a good measure of drought, which should act as a push factor for migration. The sign for excess precipitation is uncertain ex ante because excess precipitation does not necessarily imply that a flood has occurred, since flood events also depend on topology of the land apart from the level of precipitation. Excess precipitation can be associated with better quality of land and growing conditions and hence one may even expect $a_8 < 0$ for excess precipitation, at least below extreme levels of precipitation.

### 3.2 Estimation method

The specification (2) is based on a semi log form. This represents a problem for those state pairs where the migration flows equal zero, since dropping such observations from the data set may generate selection bias. On the Indian sample such state pairs represent 10% of the total number of observations. One method to avoid sample selection problems from excluding the observations with migration equal to zero, is to add one to each bilateral migration rate observation. Nevertheless, the problem remains that the log-linear specification will cause Ordinary Least Squares

(OLS) estimation of the elasticities to be inconsistent in the presence of heteroskedasticity in the error term ($u_{ij,t}$)[3] (Santos Silva and Tenreyro, 2006). Instead Santos Silva and Tenreyro (2006) demonstrate that a Poisson pseudo maximum likelihood estimator (PPML) with robust standard errors produces consistent estimates in a non-linear model. The assumption of equality between the standard deviation and the mean of the dependent variable that is characteristic of the standard Poisson maximum likelihood estimator (Poisson MLE) is no longer necessary in the PPML method. We thus rely on the results with the PPML estimator.

Another potential source of bias has been labelled multilateral resistance in the application of gravity models (Anderson, 2011). It implies that the bilateral migration rate would depend not only on the comparison between the origin and the destination state characteristics, but also on the opportunities in all the alternative destinations. The estimating equation is derived using the assumption that the error terms are distributed according to an extreme value type-1 distribution, effectively assuming independence from irrelevant alternatives for migration. Nevertheless, if this assumption does not hold, and if what has been called multilateral resistance should be accounted for, Feenstra (2002) suggests that including time-varying fixed effects for destination states accounts for multilateral resistance. Bertoli and Fernández-Huertas Moraga (2013) suggest using Pesaran's common correlated effects estimator but we cannot do this on our data that consists of two census rounds only, and follow Beine and Parsons (2015) and control for possible multilateral resistance through the inclusion of destination state and time fixed effects.

# 4 Data and measures of climate variability

## 4.1 Definition of migration

The definition of a migrant in the Indian Census is based on intent of staying rather than on a minimum duration of stay. In the census, migration flows are identified by the current place of residence (destination state), by the place of residence of provenance (origin state) and with different duration of stay. We use the 1 year duration in order to keep a strict causality between the push factors and the migration but also to minimize measurement error linked to subsequent moves. Our dependent variable is thus the gross migration flow $m_{ij,t}$ from state $i$ to state $j$ between year $t-1$ and year $t$, divided by the population that did not move in the same period, and multiplied by 100,000 for scaling purposes.

## 4.2 Climate variability: The Standardized Precipitation Index (SPI)

Rainfall is the main factor of vulnerability to water availability. The scarcity of water has negative consequences on food availability and human health and may cause diseases and displacement of populations (IPCC, 2014). The consequences in urban areas can be the difficulty to cover the requirements in drinking water in quantity as well as in quality. In rural areas, output and quality of the crops are affected in addition. The agricultural sector in India is particularly vul-

[3] The Breusch-Pagan/Cook-Weisberg test on heteroskedasticity in an OLS regression on the data leads to a test statistic of 365.37 and a $p$-value of 0. The null hypothesis of homoskedasticity is thus rejected.

nerable to water availability (O'Brien et al., 2004). To test the hypothesis that climate variability and adverse weather events act as push factors for internal migration, we compute normalized measures of low precipitation (”droughts”) and excess precipitation (”floods”) using precipitation data from the Climatic Research Unit (CRU) of the University of East Anglia. The CRU data was constructed by assimilating the observations from meteorological stations across the world in 0.5 degrees latitude by 0.5 degrees longitude grids covering the land surface of the earth (for more details see Harris et al., 2014). We use the CRU TS3.21 dataset mapped to district by month from 1901 to 2012.

From the precipitation data, we calculate the SPI, a frequently used standardized measure of drought developed by McKee et al. (1993). Conceptually, the SPI represents a z-score or the number of standard deviations above or below that an event is from the mean, for which the mean and the standard deviation are calculated over past periods (here 1901 to 2001) by fitting a gamma distribution over long run precipitation data. By using the SPI we can determine drought or excess precipitation (”flood”) for a period in a given place.

The main advantages of this measure is that it takes into account the space and temporal deviation and that it gives us a measure of the start, length and intensity of a drought or a period with excess precipitation, rather than only the absolute value of precipitation and temperature. Additionally, it allows us to have a measure with a fixed mean and variance, which makes the SPI of different locations comparable. While the SPI was developed as a drought measure it has also been suggested to be a good indicator of flood (see for instance Seiler et al., 2002) but actual floods depend not only on the quantity of rainfall but also on the soil of floodbanks and the topology of the landscape. We thus do not interpret positive deviations of the SPI as necessarily implying a flood, but prefer to refer to excess precipitation.

The raw data are on a district level and to aggregate the data on a state level, we calculate the average of the SPI in every state (a principal component analysis is presented in Appendix B as a test of this procedure). We create three variables based on the SPI to measure the frequency, the duration and the magnitude of drought and excess precipitation:

1. *Frequency*: This is a binary variable by state that takes the value of 1 if moderate or severe drought/excess precipitation was recorded in a month in that state, and 0 otherwise. The frequency measure is the number of months with drought/excess precipitation in the origin state during the five years preceding migration, to account for persistence in the effects of drought/excess precipitation.[4] The measures count total months of either severe or moderate drought/excess precipitation, but extreme events are not common on the state level data. Aggregation at a state level takes out any extreme events at a finer district-level and may lead to less precise results. More frequent drought/excess precipitation increase expectations of future similar events, and thus higher frequency should encourage migration.

2. *Maximal duration*: In the aim to catch the impact of a long period of drought or excess precipitation, we compute the maximal number of months that such an event lasted in the

[4]Barrios et al. (2006) and Strobl and Valfort (2013) also use a lags of five years for the impact of natural disasters and climate variables. Estimations in Table C.4 in Appendix C show the results with different lags.

five years preceding migration. A long duration of drought or excess precipitation in a given period is more likely to have a strong negative impact on livelihoods and hence encourage migration in search for better economic conditions.

3. *Magnitude*: This variable is defined as the sum of the absolute values of the SPI for drought or excess precipitation in the five years preceding migration. Severe or extreme drought/excess precipitation can affect people by destroying their crops or capital, as well as causing injuries and so encourage or even force migration.

These are widely used measures of climate variability and two main dimensions of drought or excess precipitation (Zargar et al., 2011). Above all, these measures are strictly exogenous and not influenced by economic activity at the time-scale considered here. We also constructed and tested measures that take into account interaction effects, such as a long and severe drought, but they were never significant and we do not present them here.

### 4.3 Other migration determinants

Since the climatic factors are not the only determinants of migration, we control also for the most important social and economic drivers. We estimate bilateral migration rates as a function of distance, common border and common language, climate variability, other non-climate natural disasters, and violence. We also control for network factors that may reduce the bilateral cost of migration. In the Indian context, we use the ratios of scheduled castes and tribes in the destination state compared to the origin state. A detailed explanation of the measures, data sources and descriptive statistics can be found in Appendix A.

## 5 Results

### 5.1 OLS vs. PPML estimators

We first present a comparison between the OLS and PPML estimators in Table 1. We only include the standard bilateral gravity variables and drought frequency in the origin state, for comparative purposes. All estimations include origin state fixed effects and destination-time fixed effects. Columns (1) and (2) correspond to the OLS estimates with the logarithm of the bilateral migration ratio as dependent variable. Columns (3) and (4) present the PPML estimates, in this case with the bilateral migration ratio as dependent variable. In columns (2) and (4) we exclude the observations with zero migration flows (10% of the observations) and columns (1) and (3) present the estimation results using all the observations.

While the coefficients in the two PPML estimations are rather similar, in the OLS estimations (1) and (2) all the coefficients differ in their value and significance level (as found by Tenreyro, 2007, on trade data). This shows the bias introduced by the ad hoc solution of adding one to the zero migration flows and applying OLS. Heteroskedasticity can thus lead to a misinterpretation of the results, yielding different conclusions depending on the sample chosen. If we compare OLS and PPML on the same sample (columns (1) and (3) or columns (2) and (4)), the role of distance is overestimated and border and language are underestimated in OLS.

Drought frequency appears more significant in the OLS estimations than in the PPML estimations with an over or underestimation depending on the sample. The results vary much between OLS and PPML estimations and the conclusions based on standard OLS estimations could thus induce the wrong conclusion on the role of climate variability. From now on, we use PPML accounting for zero observations for all estimations and present only those results.

Table 1: Inter-state migration and drought frequency: Comparison between OLS and PPML

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Estimator | OLS | OLS | PPML | PPML |
| Dependent variable | $\ln(\frac{m_{ij,t}}{pop_{ii,t}}+1)$ | $\ln(\frac{m_{ij,t}}{pop_{ii,t}})$ | $\frac{m_{ij,t}}{pop_{ii,t}}$ | $\frac{m_{ij,t}}{pop_{ii,t}} > 0$ |
| ln distance$_{ij}$ | -1.940*** | -1.107*** | -0.678*** | -0.598*** |
| | (0.104) | (0.050) | (0.078) | (0.076) |
| border$_{ij}$ | 0.394** | 1.136*** | 1.219*** | 1.276*** |
| | (0.154) | (0.087) | (0.147) | (0.140) |
| language$_{ij}$ | -0.196 | 0.165* | 0.403** | 0.347** |
| | (0.195) | (0.090) | (0.159) | (0.142) |
| drought frequency$_{it}$ | 0.017** | 0.010** | 0.014* | 0.014** |
| | (0.008) | (0.003) | (0.008) | (0.007) |
| Origin-state FE | Yes | Yes | Yes | Yes |
| Destination-state/time FE | Yes | Yes | Yes | Yes |
| N | 1860 | 1673 | 1860 | 1673 |
| $R^2$ | 0.618 | 0.848 | 0.696 | 0.726 |

Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

## 5.2 Migration, drought and excess precipitation

Climate variability can have both a direct and an indirect effect on internal migration, notably via income (Feng et al., 2012). Indeed, Table C.5 in Appendix C presents the results of estimations of ratio of the net state domestic product ($\frac{NSDP\_per\_capita_{jt}}{NSDP\_per\_capita_{it}}$) on drought and excess precipitation respectively. In Table C.5 the three drought variables (frequency, duration and magnitude) in the state of origin have a positive and statistically significant impact on the income ratio between the destination and origin states. More frequent, longer and severe droughts in the origin state indeed increase the difference in income between the destination and the origin states and can thus encourage migration indirectly. Flood frequency, duration and magnitude have a statistically significant negative impact on the income ratio between the destination and the origin states. This unexpected sign may indicate that our measures based on the SPI do not capture real floods, but only excess precipitation. Given that all the variables representing drought and excess precipitation in the origin state have a highly significant impact on the income ratio, we should indeed exclude the income ratio in the estimations of bilateral migration rates. By excluding income in the estimations, we thus estimate the net effect on migration, without separating direct and indirect effects.

Table 2 presents the estimations of equation 2 with the drought (columns (1)-(3)) and excess precipitation measures (columns (4)-(6)) in the origin state. We introduce the variables that measure drought or excess precipitation (frequency, the longest duration and magnitude) separately in the estimations because of the high correlation between them (see Table C.3 in Appendix C).

The results show that the proxies for the cost of migration are the most important factors for internal migration, both in value and in statistical significance. Bilateral migration rates between contiguous states are 2.4 times larger than for states that do not share a common border. States that share a common language have 50% larger bilateral migration rates.[5] Geographical distance is also statistically significant with a 1% larger distance decreasing the bilateral migration rate by 0.7%. The differences in scheduled caste and scheduled tribe rates between the destination and the origin state are not significant. The per capita homicide rate is not significant either, but has the positive sign predicted ex ante according to the push factor hypothesis concerning criminality rate.

Earthquakes in the origin state have a positive effect on bilateral migration rates at a 10% significance level in estimations (1), (3) and (6), but the magnitude and the sign of the coefficients remain stable, with a marginal effect on out-migration of approximately a 60% increase in the migration rate for each additional earthquake. Among the three drought measures tested, the role of push factor for migration is rejected for the duration of the longest drought. The results for the variability measure - the frequency of drought events compared to the long run mean - indicate that an additional month of drought during the five years preceding migration would increase the bilateral migration rate by 1.6% (column (1)) and one additional unit increase in the SPI in absolute magnitude (which is very high) will increase the migration rate by 0.9% but the evidence of this effect is weak since it is only significant at a 10% level.

The coefficients of the corresponding measures of excess precipitation are negative, but with weak statistical significance for frequency and magnitude. We thus conclude that when precipitation variability is significant as a push factor for inter-state migration in India it is mainly through drought. This result may be explained by several factors. First, on the sample studied here there is less variability in excess precipitation than in drought between 1991 and 2001 (as seen in Figure 3 in Appendix A). Second, the measures we use are based on the SPI, which is a reliable drought indicator, but a less direct measure of flood, since it captures the hydrological and climatological conditions for floods but not other factors, such as topology. Guiteras et al. (2015) compare precipitation data with remote sensing data on actual flooding in Bangladesh and argue that precipitation data are only a weak proxy for floods. We address this concern in section 5.3.1 where we test alternative measures of flood. Third, existing evidence from other countries, notably Bangladesh (Gray and Mueller, 2012) shows that floods do not always induce migration. Drought can be characterized as a long-run process, that does not always induce an immediate response, but when it does it can lead to permanent migration. Flooding is a rapid on-

[5]The marginal effects for dummy variables are calculated as $(e^{b_i} - 1)$ where $b_i$ is the estimated coefficient of the variable.

set phenomenon, to the contrary, that may lead to short-distance displacement only (Barnett and Webber, 2010; Piguet, 2010). Responses to flood events are therefore different, and if migration occurs, it may be temporary rather (Perch-Nielsen et al., 2008).

Table 2: Inter-state migration, drought and excess precipitation

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| ln distance$_{ij}$ | -0.677*** | -0.677*** | -0.678*** | -0.679*** | -0.677*** | -0.680*** |
| | (0.079) | (0.078) | (0.078) | (0.077) | (0.077) | (0.077) |
| border$_{ij}$ | 1.223*** | 1.218*** | 1.220*** | 1.219*** | 1.216*** | 1.220*** |
| | (0.149) | (0.148) | (0.148) | (0.147) | (0.148) | (0.148) |
| language$_{ij}$ | 0.400** | 0.399** | 0.398** | 0.400** | 0.399** | 0.399** |
| | (0.159) | (0.160) | (0.160) | (0.159) | (0.161) | (0.160) |
| $\ln \frac{SC_{jt}+1}{SC_{it}+1}$ | 1.590 | -4.869 | -3.695 | 13.544 | -1.458 | 12.879 |
| | (20.892) | (20.935) | (21.188) | (22.023) | (20.789) | (22.312) |
| $\ln \frac{ST_{jt}+1}{ST_{it}+1}$ | -2.498 | -3.392 | -2.633 | -4.227 | -3.096 | -3.816 |
| | (6.566) | (6.469) | (6.489) | (6.426) | (6.499) | (6.436) |
| ln murder pc$_{it}$ | 0.387 | 0.308 | 0.367 | 0.346 | 0.324 | 0.407 |
| | (0.287) | (0.286) | (0.288) | (0.287) | (0.285) | (0.287) |
| earthquake$_{it}$ | 0.636* | 0.561 | 0.681* | 0.532 | 0.564 | 0.576* |
| | (0.352) | (0.344) | (0.356) | (0.344) | (0.347) | (0.344) |
| drought frequency$_{it}$ | 0.016** | | | | | |
| | (0.008) | | | | | |
| longest drought dur$_{it}$ | | 0.010 | | | | |
| | | (0.007) | | | | |
| drought magnitude$_{it}$ | | | 0.009* | | | |
| | | | (0.005) | | | |
| flood frequency$_{it}$ | | | | -0.014* | | |
| | | | | (0.007) | | |
| longest flood dur$_{it}$ | | | | | -0.002 | |
| | | | | | (0.007) | |
| flood magnitude$_{it}$ | | | | | | -0.010* |
| | | | | | | (0.005) |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination-state/time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 |
| $R^2$ | 0.696 | 0.694 | 0.693 | 0.697 | 0.690 | 0.696 |

The dependent variable is the bilateral migration rate from state $i$ to state $j$ between year $t-1$ and year $t$. Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

## 5.3 Robustness tests

### 5.3.1 Alternative climatic measures and irrigation

A potential problem could be omitted variable bias resulting from only including precipitation data and not temperature (Auffhammer et al., 2013). We redo the estimations with alternative definitions of anomalies in temperature and precipitation defined as in Marchiori et al. (2012) and Beine and Parsons (2015). The variables are constructed as follows:

$$WeatherAnomaly_{i,t} = \frac{Weather_{i,t} - \mu_i^{LR}(Weather)}{\sigma_i^{LR}(Weather)}$$

| | |
|---|---|
| $LR$: | Long run period from 1901 to 2000. |
| $WeatherAnomaly_{i,t}$: | Temperature or precipitation long run deviation in state $i$ at time $t$. |
| $Weather_{i,t}$: | Temperature or precipitation level in state $i$ at time $t$. |
| $\mu_i^{LR}(Weather)$: | Temperature or precipitation long run mean for state $i$. |
| $\sigma_i^{LR}(Weather)$: | Temperature or precipitation long run standard deviation for state $i$. |

In Table 3, column (1) shows the same estimations as in Table 2 with drought frequency. In column (2), we add temperature anomalies, in order to control also for temperature anomalies and not only precipitation anomalies as done with the SPI based variables. All the coefficients - including drought frequency - remain stable, but temperature anomalies are not statistically significant. Column (3) presents the estimation with temperature anomalies only. The coefficient for temperature anomalies is positive and not significant. In column (4) precipitation anomalies replace the temperature variable. The coefficient is negative (as for our excess precipitation variables) but not statistically significant. When combining temperature and precipitation anomalies (column (5)), both variables are non significant. These results indicate that temperature anomalies measured in this manner do not play an important role in inter-state migration in India and support the relevance of the definition of the drought variables used in this analysis instead. The results can also be interpreted as evidence that other work that use such anomaly measures of climate variability for the impact on migration and that have found no evidence of an impact will maybe find different results with a better measure of variability.

In Table 4, additional climate variability measures are tested. Columns (1)-(3) present the continuous values of the SPI directly. In column (1) the measure is the absolute value of the annual SPI average 5 years before migration. This measure aims at capturing average deviations (positive or negative) in climate variability. It measures climate variability without distinguishing between positive and negative shocks, but the large size of most Indian states limits the measure. Indeed, we do not find any statistical significance. Column (2) and (3) measure only the positive or negative average of SPI superior to 1 in absolute values in order to capture negative (drought) or positive (excess precipitation) shocks separately. Again we can find a statistically significant positive impact of drought but the excess precipitation measure based on the continuous value of the SPI is negative and not significant, as before.

In column (4)-(6), we test three types of alternative flood measures based on the Dartmouth Flood Observatory data, as used in Ghimire and Ferreira (2016). The Observatory archives every large flood observed (see definition in Table A.2 in Appendix A). In Table 4 all the measures are based on floods occurring one year before migration, and we define the variables to be comparable with the excess precipitation frequency measures used previously. The flood frequency variable measures the number of months with a large flood event, flood severity measures the average flood severity index defined by the Observatory and finally, flood magnitude is the log of the product between frequency and duration. Even if the signs of the flood variables are positive as expected ex ante (contrary to the SPI based measures), none of the coefficients are statistically significant.

Table 3: Inter-state migration and long run anomalies in temperature and precipitation

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| ln distance$_{ij}$ | -0.677*** | -0.677*** | -0.677*** | -0.678*** | -0.678*** |
| | (0.079) | (0.079) | (0.077) | (0.078) | (0.078) |
| border$_{ij}$ | 1.223*** | 1.223*** | 1.216*** | 1.217*** | 1.217*** |
| | (0.149) | (0.148) | (0.148) | (0.148) | (0.148) |
| language$_{ij}$ | 0.400** | 0.401** | 0.400** | 0.400** | 0.399** |
| | (0.159) | (0.159) | (0.162) | (0.160) | (0.160) |
| $\ln \frac{SC_{jt}+1}{SC_{it}+1}$ | 1.590 | 0.762 | -1.082 | -3.129 | -2.607 |
| | (20.892) | (20.908) | (20.740) | (20.814) | (20.817) |
| $\ln \frac{ST_{jt}+1}{ST_{it}+1}$ | -2.498 | -2.504 | -2.851 | -3.195 | -3.192 |
| | (6.566) | (6.570) | (6.453) | (6.427) | (6.426) |
| ln murder pc$_{it}$ | 0.387 | 0.347 | 0.335 | 0.300 | 0.325 |
| | (0.287) | (0.291) | (0.290) | (0.291) | (0.294) |
| earthquake$_{it}$ | 0.636* | 0.587* | 0.586* | 0.675* | 0.713* |
| | (0.352) | (0.355) | (0.352) | (0.355) | (0.371) |
| drought frequency$_{it}$ | 0.016** | 0.017** | | | |
| | (0.008) | (0.008) | | | |
| temperature anomaly$_{it}$ | | -0.601 | 0.117 | | 0.351 |
| | | (0.653) | (0.614) | | (0.647) |
| precipitation anomaly$_{it}$ | | | | -2.009 | -2.141 |
| | | | | (1.300) | (1.353) |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes |
| Destination-state /time FE | Yes | Yes | Yes | Yes | Yes |
| N | 1860 | 1860 | 1860 | 1860 | 1860 |
| $R^2$ | 0.696 | 0.697 | 0.690 | 0.692 | 0.692 |

The dependent variable is the bilateral migration rate from state $i$ to state $j$ between year $t-1$ and year $t$. Robust standard errors in parentheses.
$*p < 0.10, **p < 0.05, ***p < 0.001$.

Table 4: Inter-state migration and other alternative climate variables

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| ln distance$_{ij}$ | -0.677*** | -0.677*** | -0.677*** | -0.677*** | -0.677*** | -0.677*** |
| | (0.077) | (0.077) | (0.078) | (0.077) | (0.077) | (0.077) |
| border$_{ij}$ | 1.217*** | 1.218*** | 1.219*** | 1.214*** | 1.215*** | 1.215*** |
| | (0.148) | (0.148) | (0.148) | (0.147) | (0.147) | (0.147) |
| language$_{ij}$ | 0.399** | 0.400** | 0.398** | 0.402** | 0.400** | 0.400** |
| | (0.162) | (0.162) | (0.160) | (0.160) | (0.161) | (0.161) |
| ln $\frac{SC_{jt}+1}{SC_{it}+1}$ | -2.247 | -0.823 | -5.638 | -2.264 | -1.525 | -1.482 |
| | (20.580) | (20.848) | (20.937) | (20.855) | (20.739) | (20.731) |
| ln $\frac{ST_{jt}+1}{ST_{it}+1}$ | -2.978 | -3.513 | -3.860 | -2.247 | -2.324 | -2.082 |
| | (6.432) | (6.416) | (6.469) | (6.438) | (6.456) | (6.473) |
| ln murder pc$_{it}$ | 0.350 | 0.403 | 0.327 | 0.329 | 0.332 | 0.334 |
| | (0.300) | (0.270) | (0.284) | (0.284) | (0.284) | (0.285) |
| earthquake$_{it}$ | 0.543 | 0.679* | 0.657* | 0.766** | 0.600* | 0.591* |
| | (0.367) | (0.361) | (0.350) | (0.362) | (0.341) | (0.341) |
| avg. SPI$_{it}$ | -0.070 | | | | | |
| | (0.313) | | | | | |
| avg. SPI$_{it} > 1$ | | -0.084 | | | | |
| | | (0.136) | | | | |
| avg. SPI$_{it} < 1$ | | | 0.188* | | | |
| | | | (0.108) | | | |
| flood frequency$_{it}$ | | | | 0.163 | | |
| | | | | (0.110) | | |
| flood severity$_{it}$ | | | | | 0.125 | |
| | | | | | (0.107) | |
| flood magnitude$_{it}$ | | | | | | 0.058 |
| | | | | | | (0.043) |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination-state FE | Yes | Yes | Yes | No | No | No |
| Destination-state/time FE | No | No | No | Yes | Yes | Yes |
| Bilateral FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | No | No | No |
| N | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 |
| $R^2$ | 0.690 | 0.690 | 0.694 | 0.691 | 0.690 | 0.691 |

The dependent variable is the bilateral migration rate from state $i$ to state $j$ between year $t-1$ and year $t$. Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

Adaptation measures other than migration can limit the impact of climate variability (Barnett and Webber, 2010; Mendelsohn, 2012). In agriculture, farmers can adapt to shortfalls in precipitation, in particular, or increased variability in precipitation by changing to more resistant crops or by investing in irrigation infrastructure (O'Brien et al., 2004). The analysis here is on macro level inter-state flows and we cannot control for drought resistant crops, for example. As an additional robustness test we control for irrigation capacity as one of the most common adaptation measures against drought. In particular, Taraz (2015) finds that Indian farmers adjust their irrigation investment according to monsoon rainfall variability, but that it has limited efficacy in reducing the losses in agricultural profits. The variable used is the ratio of land with net irrigation on total cultivated land in the origin state. Table C.6 in Appendix C shows the net effect of drought frequency including interaction terms with the net irrigation rate. The effect of drought frequency is lower (0.013) and only weakly significant at the 10% level. Since irrigation is correlated with climate (see Table C.3) and it is not a determinant of migration on its own, including it will only reduce the precision of the estimated coefficient on drought because of collinearity. In fact, we note that the significance and measure of the coefficient on drought is attenuated when net irrigation measures are included. Nevertheless, the effect of drought maintains its sign and order of magnitude, which confirms the robustness of its effect.

#### 5.3.2 Additional fixed effect specifications

The estimations in Table 5 aim at presenting a more complete model of migration mainly by adding bilateral fixed effects invariant in time. These dummy variables control for all the factors inducing migration that are common to origin and destination states and that remain fixed in time like the control variables included in our original estimations (common language, common border and distance between states), but also all other potential bilateral fixed effects. Indeed, we can observe that the part of the variation explained by the model moves from 70% in previous estimations to almost 95% by adding bilateral fixed effects. Columns (1)-(3) include also origin and destination state fixed effects and time fixed effects, and columns (4)-(6) origin state fixed effects and destination state/time fixed effects (as in all previous estimations).

Although the magnitude of the coefficients does not change much, the statistical significance is higher. Indeed, the murder rate coefficient is significant at the 5% level, with a 1% increase in the average mortality rate in the origin state 5 years before migration increasing the migration rate by 0.3-0.5%. Earthquakes again have a very large effect with an additional earthquake leading to a 60% increase in the migration rate. Nevertheless, this effect should be taken with caution since at maximum only one event is registered in our sample. The scheduled caste (SC) ratio is also significant in the estimations with destination and time fixed effects. Even if the magnitude of the effect of the SC and ST ratios is not stable, the negative association with migration seems stable. This confirms the intuition that larger similarity in castes between origin and destination states is likely to increase migration between these states, which confirms a network effect.

Concerning the drought measures, all the three variables are significant at the 1% to 5% level and maintain the positive sign. Nevertheless, the magnitude of the drought frequency coefficient is smaller. We should also take into account that the PPML estimator has excluded 86 observations in the sample in order to converge.[6] As an additional robustness test we thus run OLS regressions

[6] In fact, the PPML estimator may have problems to converge with too many dependent variables equal to zero or with too many dummies. Since we have only 10% of the observations with zero migration rate and since we do

which include the total observations (see Table C.7 in Appendix C). The OLS results show that, even if the coefficient magnitudes are larger, the impact of drought on migration is always robust.

Table 5: Inter-state migration and drought with bilateral fixed effects

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| $\ln \frac{SC_{jt}+1}{SC_{it}+1}$ | -9.059** | -11.186** | -10.270** | 2.268 | -2.571 | -1.032 |
| | (4.126) | (4.014) | (4.231) | (5.984) | (5.282) | (6.292) |
| $\ln \frac{ST_{jt}+1}{ST_{it}+1}$ | -1.638 | -2.150 | -1.754 | -2.772 | -3.401* | -2.762 |
| | (2.202) | (2.082) | (2.047) | (2.183) | (2.065) | (2.065) |
| ln murder $\text{pc}_{it}$ | 0.503** | 0.404** | 0.440** | 0.321** | 0.247** | 0.301** |
| | (0.155) | (0.154) | (0.152) | (0.116) | (0.114) | (0.117) |
| $\text{earthquake}_{it}$ | 0.652*** | 0.588*** | 0.657*** | 0.590*** | 0.524*** | 0.625*** |
| | (0.069) | (0.059) | (0.068) | (0.129) | (0.115) | (0.118) |
| drought $\text{frequency}_{it}$ | 0.010** | | | 0.013*** | | |
| | (0.004) | | | (0.003) | | |
| longest drought $\text{dur}_{it}$ | | 0.010** | | | 0.009** | |
| | | (0.004) | | | (0.003) | |
| drought $\text{magnitude}_{it}$ | | | 0.006** | | | 0.007*** |
| | | | (0.002) | | | (0.002) |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination-state FE | Yes | Yes | Yes | No | No | No |
| Destination-state/time FE | No | No | No | Yes | Yes | Yes |
| Bilateral FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | No | No | No |
| N | 1774 | 1774 | 1774 | 1774 | 1774 | 1774 |
| $R^2$ | 0.939 | 0.941 | 0.936 | 0.968 | 0.968 | 0.966 |

The dependent variable is the bilateral migration rate from state $i$ to state $j$ between year $t-1$ and year $t$. Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

### 5.3.3 Male migration rates, drought and excess precipitation

The Indian Census incorporates a question on the reason for migration, with the possible answers being work/employment, business, education, marriage, moved after birth, moved with household and other. As shown in Table C.2 in the Appendix C, family moving was cited as the predominant reason for migration among women (41% of women in 1991 and 48% of women in 2001) and employment for men (42% of men in 1991 and 54% of men in 2001). To further test the relationship between climate variability and internal migration in India, we did separate estimations on male migration rates, that constitute 52% in 1991 and 57% in 2001 of total migrants. The

not encounter problems with convergence in the other estimations, we argue that the convergence problem here is due to the quantity of dummy variables that could lead to collinearity problems. The PPML command in Stata is able to first identify and drop problematic regressors and then run the usual PPML estimation. We verified the observations dropped and there do not seem to be "special pairs" of bilateral migration that are systematically dropped. Selection bias does not seem to be a problem in these estimations.

estimations in Table 6 show that the coefficients are of about the same size and significance as in the estimations on the total migration rates in Table 2. Most importantly, drought frequency, duration and magnitude have the same marginal effect and significance level.

Table 6: Inter-state male migration and drought

| | (1) | (2) | (3) |
|---|---|---|---|
| ln distance$_{ij}$ | -0.699*** | -0.699*** | -0.700*** |
| | (0.082) | (0.081) | (0.082) |
| border$_{ij}$ | 1.100*** | 1.095*** | 1.097*** |
| | (0.151) | (0.151) | (0.150) |
| language$_{ij}$ | 0.487** | 0.487** | 0.486** |
| | (0.162) | (0.163) | (0.164) |
| ln $\frac{SC_{jt}+1}{SC_{it}+1}$ | 2.441 | -3.782 | -2.504 |
| | (21.660) | (21.797) | (22.030) |
| ln $\frac{ST_{jt}+1}{ST_{it}+1}$ | -2.223 | -3.077 | -2.311 |
| | (6.651) | (6.601) | (6.615) |
| ln murder pc$_{it}$ | 0.442 | 0.362 | 0.426 |
| | (0.310) | (0.308) | (0.311) |
| earthquake$_{it}$ | 0.624* | 0.554 | 0.675* |
| | (0.349) | (0.342) | (0.353) |
| drought frequency$_{it}$ | 0.015** | | |
| | (0.007) | | |
| longest drought dur$_{it}$ | | 0.010 | |
| | | (0.007) | |
| drought magnitude$_{it}$ | | | 0.009* |
| | | | (0.005) |
| Origin-state FE | Yes | Yes | Yes |
| Destination-state/time FE | Yes | Yes | Yes |
| N | 1860 | 1860 | 1860 |
| $R^2$ | 0.673 | 0.669 | 0.668 |

The dependent variable is the bilateral migration rate from state $i$ to state $j$ between year $t-1$. Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

## 6 Conclusions

The objective of the paper is to test the hypothesis that climate variability acts as a push-factor on internal migration. We estimate Indian bilateral inter-state migration rates calculated using census data from 1991 and 2001 on variability measures based on deviations from the long run climate over the period 1901-2000. To the best of our knowledge this is one of few attempts to investigate the impact of climate variability on internal migration using the precise and complete census data on the level of such a large and diverse country as India. The use of migration flow data defined between years $t-1$ and $t$ enables us to match the data more precisely in time with the potential determinants of migration, in particular the climatic factors, other natural disasters, and socio-economic characteristics of the origin and destination rates. This is an advantage compared to the existing literature on climatic factors and migration, that often analyses average figures over a longer time period. The other main contribution of the analysis is to use relevant objective meteorological indicators of climate variability, based on the standardized precipitation index. We create three variables based on the SPI to distinguish between the variability in precipitation, on the one hand, and two important dimensions of deviations in precipitations - the duration and the magnitude of drought and excess precipitation. The analysis furthermore controls for the econometric problems that arise when applying a gravity-type model on bilateral migration flows. In particular, we apply the Poisson pseudo maximum likelihood estimator to correct for the presence of zero migration flows between certain states and avoid heteroskedasticity.

The estimation results show significant effects on bilateral migration rates from the frequency of droughts. An additional month of drought increases the bilateral inter-state migration rate by 1.6%, all else equal. We suggest that the findings may be interpreted as evidence of the expectations of future drought inducing migration. Observed frequency of droughts tends to reinforce future expectations of drought and may hence induce migration. The effect seems small, though, especially when compared to the important role of barriers to migration in the Indian context that explain the low Indian inter-state migration rates. Nevertheless, the results show that an increase in climate variability measured as drought frequency can induce additional large numbers of inter-state migrants in absolute values. We do not find a consistently significant effect of the magnitude of the droughts that occurred in the period preceding migration nor of the duration of the longest drought. There is no evidence of higher frequency, duration or intensity of excess precipitation acting as push factors for inter-state migration.

Several reasons may explain why we find only a significant effect from drought and not from excess precipitation on bilateral inter-state migration rates in India. In the data over the time of the two censuses studied here (1991 and 2001) there was less variability in excess precipitation and, in particular, the excess precipitation was never extreme with respect to the long run standardized measures and geographical zones used here (states). Alternative measures from the Dartmouth Flood Observatory were never significant either, although showing the expected positive sign on migration. An extension of the analysis using more exact measures of flood, by for example exploiting remote sensing data as in Guiteras et al. (2015) or in Gröger and Zylberberg (2016), seems a useful direction for future research. Another limitation of the current analysis is the lack of detailed data on other natural disasters at a state level. We also believe more disaggregated analysis would be useful, but the district data do not give the origin district of the migrants, hence bilateral migration rates could not be calculated.

## References


Anderson, J.E. (2011). The gravity model. *Annual Review of Economics* 3, 133-160.

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

Attri, S.D. & Tyagi, A. (2010). Climate profile of India. Government of India Ministry of Earth Sciences India Meteorological Department. New Delhi.

Auffhammer, M., Hsiang, S., Schlenker, W. and Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy* 7(2), 181-198.

Barnett, J. & Webber, M. (2010). Migration as adaptation: Opportunities and limits. In McAdam, J. (Ed.) *Climate change and displacement: Multidisciplinary perspectives*, pp. 37-55. Oxford: Hart Publishing.

Barrios, S., Bertinelli L. & Strobl, E. A. (2006). Climatic change and rural-urban migration: The case of sub-Saharan Africa. *Journal of Urban Economics* 60(3), 357-371.

Barrios, S., Bertinelli L. & Strobl, E. A. (2010). Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy. *The Review of Economics and Statistics* 92(2), 350-366.

Beine, M. & Parsons, C. (2015). Climatic factors as determinants of international migration. *Scandinavian Journal of Economics* 117(2), 723-767.

Beine, M., Docquier, F. & Özden, C. (2011). Diasporas. *Journal of Development Economics* 95(1), 30-41.

Bertoli, S. & Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics* 102, 79-100.

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

Bodvarsson, Ö. & Van den Berg, H. (2009). *The economics of immigration*. Berlin Heidelberg: Springer-Verlag.

Bohra-Mishra P., Oppenheimer M. & Hsiang S. (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences of the United States* 111(27), 9780-85.

Boustan L.P., Kahn, M. & Rhode PW. (2012). Moving to higher ground: Migration response to natural disasters in the early twentieth century. *American Economic Review* 102(3), 238-44.

Burke, M., Hsiang, S.M. & Miguel E. (2015). Global non-linear effect of temperature on economic production. *Nature* 527, 235-239.

Census of India (2011). Provisional Population Totals. Office of the Registrar General and Census Commissioner, India. Paper 1 of 2011, India Series 1.

Coniglio, N. & Pesce, G. (2015). Climate variability and international migration: An empirical analysis. *Environment and Development Economics* 20(4), 434-468.

Czaika, M. (2011). Internal and international migration as a response to double deprivation: Some evidence from India. IMI Working Papers Series No. 37

DeGaetano, A. T. (2001). Spatial grouping of United States climate stations using a hybrid clustering approach. *International Journal of Climatology* 21(7), 791-807.

Dell, M., Jones, B.F. & Olken, B.A. (2009). Temperature and income: Reconciling new cross-sectional and panel estimates. *American Economic Review, Papers and Proceedings* 99(2), 198-204.

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

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

Feenstra, R.C. (2002). Border effects and the gravity equation: Consistent methods for estimation. *Scottish Journal of Political Economy* 49(5), 491-506.

Feng, S., Oppenheimer, M. & Schlenker, W. (2012). Climate change, crop yields and internal migration in the United States. NBER Working Paper 17734.

Gemenne F. (2011). Why the numbers do not add up: A review of estimates and predictions of people displaced by environmental changes. *Global Environmental Change* 21S, S41-S49.

Ghimire, R. & Ferreira, S. (2016). Floods and armed conflict. *Environment and Development Economics*, forthcoming.

The Government Office for Science, London (2011). *Foresight: Migration and global environmental change*, Final Project Report.

Gray, C. & Mueller, V. (2012). Natural disasters and population mobility in Bangladesh. *Proceedings of the National Academy of Sciences* 109(16), 6000-6005.

Gröger, A. & Zylberberg, Y. (2016). Internal labour migration as a shock coping strategy: Evidence from a typhoon. *American Economic Journal: Applied Economics*, forthcoming.

Guerreiro, M., Lajihna, T. & Abreu I. (2008). Flood analysis with the standardized precipitation index. *Revista da Faculdade de Ciencia e Tecnologia de la Universidade Fernando Pessoa* 4, 8-14.

Guiteras, R., Jina, A. & Mobarak, A.M. (2015). Satellites, self-reports, and submersion: Exposure to floods in Bangladesh. *American Economic Review: Papers and Proceedings* 105(5), 232-236.

Halliday T. (2006). Migration, risk and liquidity constraints in El Salvador. *Economic Development and Cultural Change* 54(4), 893-925.

Halliday T. (2012). Intra-household labor supply, migration, and subsistence constraints in a risky environment: Evidence from rural El Salvador. *European Economic Review* 56(6), 1001-1019.

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H. (2014) Updated high resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology* 34(3), 623-642.

Intergovernmental Panel on Climate Change (IPCC) (2014). Climate Change 2014: Impacts, Adaptation, and Vulnerability, The Working Group II contribution to the IPCC Fifth Assessment Report (WGII AR5).

Karemera, D., Oguledo, V. & Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied Economics* 32(13), 1745-1755.

Kumar, K.S.K. & Viswanathan, B. (2013). Influence of weather on temporary and permanent migration in rural India. *Climate Change Economics* 4(02), 1-19.

Lilleor, H.B. & Van den Broeck, K. (2011). Economic drivers of migration and climate change in LDCs. *Global Environmental Change* 21S, S70-S81.

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

Mayda, A. M. (2010). International migration : A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics* 23(4), 1249-1274.

McKee, T., Doesken, N. & 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.

Mendelsohn, R. (2012). The economics of adaptation to climate change in developing countries. *Climate Change Economics* 3(2), 1250006, 1-21.

Mitra, A. & Murayama, M. (2008). Rural to urban migration: A district level analysis for India. IDE Discussion Paper. No. 137.

Mitra, A. & Ray, D. (2014). Implications of an economic theory of conflict: Hindu-muslim violence in India. *Journal of Political Economy* 122(4), 719-765.

Munoz-Diaz, D. & Rodrigo, F. (2004). Spatio-temporal patterns of seasonal rainfall in Spain (1912-2000) using cluster and principal component analysis: Comparison. *Annales Geophysicae* 22, 1435-1448.

Munshi, K. & Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration and the rural-urban wage gap. *American Economic Review* 106(1), 46-98.

Myers, N. (1997). Environmental refugees. *Population and Environment* 19(2), 167-182.

O'Brien, K., Leichenko, R., Kelkar, U., Venema, H., Aandahl, G., Tompkins, H., Javed, A., Bhadwal, S., Barg, S., Nygaard, L. & West, J. (2004). Mapping vulnerability to multiple stressors: Climate change and globalization in India. *Global Environmental Change* 14, 303-313.

Özden, Ç. & Sewadeh, M. (2010). How important is migration? In E. Ghani, ed., *The poor half million in South Asia: What is holding back lagging regions?*, pp. 294-322. New Delhi, India: Oxford University Press.

Perch-Nielsen, S.L., Bättig, M.B. & Imboden, D. (2008). Exploring the link between climate change and migration. *Climatic Change* 91, 375-393.

Piguet E. (2010) Linking climate change, environmental degradation and migration: A methodological overview. *Climate Change* 1(4), 517-524.

Reuveny, R. & Moore, W. (2009). Does environmental degradation influence migration? Emigration to developed countries in the late 1980s and 1990s. *Social Science Quarterly* 90(3), 461-479.

Santos Silva, J.M.C. & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics* 88(4), 641-658.

Seiler, R.A., Hayes, M. & Bressan, L. (2002). Using the Standardized Precipitation Index for flood risk monitoring. *International Journal of Climatology* 22, 1365-1376.

Singh, K., & Singh, H. (1996). Space-time variation and regionalization of seasonal and monthly summer monsoon rainfall of the sub-Himalayan region and Gangetic plains of India. *Climate Research* 6, 251-262.

Stern, N. (2007). *The economics of climate change: the Stern review.* Cambridge: Cambridge University Press.

Strobl, E.A. & Valfort, M. (2013). The effect of weather-induced internal migration on local labor markets: Evidence from Uganda. *World Bank Economic Review* 29(2), 385-412.

Taraz, V. (2015). Adaptation to climate change: Historical evidence from the Indian monsoon. Accessed from http://www.smith.edu/economics/VisTaraz.php.

Tenreyro, S. (2007). On the trade impact of nominal exchange rate volatility. *Journal of Development Economics* 82, 485-508.

Van Lottum, G. & Marks, D. (2010). The determinants of internal migration in a developing country: Quantitative evidence for Indonesia, 1930-2000. *Applied Economics* 44(34), 4485-4494.

Viswanathan, B. & Kumar, K.S.K. (2015). Weather, agriculture and rural migration: Evidence from state and district level migration in India. *Environment and Development Economics* 20(4), 469-492.

Zargar, A., Sadiq, R., Naser, B. & Khan, F.I. (2011). A review of drought indices. *Environmental Reviews* 19, 333-349.

# A Detailed description of the data

## A.1 Area and period studied

We use bilateral inter-state migration data from the Indian Census of 1991 and 2001.[7] Between 1991 and 2001, India changed the territorial administrative division of its states. In 1991, India counted 27 states and 5 Union Territories. In 2001, 3 states were divided in two[8], resulting in a total of 30 states and 5 Union Territories. To unify the database, we use the territorial administrative division of 1991. Hence, for 2001, we aggregate the data of the divided states as they were defined in 1991. We analyse the Union Territories as states. Since we do not have data from 1991 on the state of Jammu and Kashmir[9], we removed this state from the sample. Jammu and Kashmir represent only 1% of the Indian population. The final sample thus counts 31 states for 1991 and 2001. As the analysis of migration is made in a bilateral manner, we have 930 observations (31x30, migration between the same states being 0) for each year.

## A.2 Net State Domestic Product (NSDP)

The NSDP per capita is used as a measure of the income per capita of the state. We use the database of the Reserve Bank of India and calculate the deflated NSDP at constant price for the two years of interest (1990 and 2000).

The variable used is the ratio of the NSDP per capita of the destination state divided by that of the origin state, in the year preceding migration ($t-1$), in order to reduce any endogeneity with the migration flows, which occur between year $t-1$ and year $t$.

## A.3 Distance between states

We calculate the distance between different states, by taking the most populated city as reference city, most often the capital of the state, but in some cases the economic center of the state, according to the great circle formula.[10]

$$d_{ij} = R * cos^{-1}(sin(a)sin(b) + cos(a)cos(b)cos(c-d)) \tag{3}$$

[7] The population census in India is conducted every ten years, but we only had access to computerized data from 1991 onwards. Data from 2011 on inter-state migration flows are not yet available.

[8] Uttar Pradesh, Bihar and Madhya Pradesh, that have given rise to the states Uttaranchal, Jharkhand and Chhattisgarh respectively.

[9] The census was not conducted in the state of Jammu and Kashmir in 1991.

[10] The latitudes and longitudes of the largest cities in every state can be found on the website "Maps of India". See www.mapsofindia.com

$d_{ij}$: distance between state $i$ and state $j$

$R$: equatorial radius, equal to 6,378 km

$a$: latitude degree of state $i$

$b$: latitude degree of state $j$

$c$: longitude degree of state $i$

$d$: longitude degree of state $j$

As explanatory variable we use the distance between two states, measured in km.

## A.4 Common border and common language

We introduce a dummy variable to control for neighbouring states. It takes the value of one for bilateral migration where the origin and destination states have a common border, and zero otherwise.

One of the specificities of India is that there are 22 different native languages (English excluded) inside the country. As another proxy of the cost of migration, we introduce a language dummy variable. It takes the value of one for bilateral migration where the origin and destination states share a common language, and zero otherwise. To assign a language to a state, we took the major language spoken in the state. The source of this variable is "Maps of India".

These two variables are proxies for cultural and traditional similarities between states.

## A.5 Standardized Precipitation Index (SPI) and alternative climatic factors

The SPI is a well-recognized measure of variability in precipitation from its long run mean.[11] It is constructed by fitting a gamma distribution on long run precipitation data, within a defined scale (here 12 months). The SPI can take values between -3 and 3 and a (moderate) drought begins when the SPI has a value of -1 (precipitation falls one standard deviation below its historical mean) and goes on in time until the SPI becomes positive again. In that way, it is possible to determine the beginning and end date and hence to calculate the length of a given drought episode. The intensity of the drought is measured according to the value of the SPI. An excess of precipitation can be measured following the same logic. It begins with a value of +1 (precipitation increases by one standard deviation above its historical mean) and continues until the SPI becomes negative. Table A.1 illustrates the definition of intensity of drought and excess precipitation with this method.[12]

We compare variability in precipitation between states by counting the number of times precipitation exceeds or is below the long run mean on a normalized scale. Figure 3 illustrates the raw data of the analysis. The figure shows the number of months with one standard deviation

[11] The Lincoln Declaration on Drought Indices (11 December 2009, Lincoln, USA) recommended that *The National Meteorological and Hydrological Services (NMHSs) around the world use the SPI to characterize meteorological droughts and provide this information on their websites, in addition to the indices currently in use. WMO was requested to take the necessary steps to implement this recommendation.*

[12] For more details on the SPI, see McKee et al. (1993)

Table A.1: Definitions of drought and excess precipitation ("flood") according to the SPI

| **SPI values** | **Category** |
|---|---|
| 0 to -0.99 | Mild drought |
| -1 to -1.49 | Moderate drought |
| -1.5 to -1.99 | Severe drought |
| <= -2 | Extreme drought |
| 0 to 0.99 | Mild flood |
| 1 to 1.49 | Moderate flood |
| 1.5 to 1.99 | Severe flood |
| >= 2 | Extreme flood |

Source: McKee et al. (1993) for drought and Guerreiro et al. (2008) for flood.

or more of either low precipitation ("drought") or excess precipitation ("flood") in the five years preceding the census in either 1991 and in 2001. The first thing to note is that the months with drought by state varied much between 1991 and 2001, whereas there is less variation over time for the number of months with excess precipitation by state. Overall, several of the states record no occurrence of drought or excess precipitation at all in the five years preceding 2001. The states with a high number of months with low precipitation in the five years preceding 1991 were Kerala and Madhya Pradesh, in addition to several small states and island states, and Bihar, Tripura and Nagaland in 2001.

The states with the highest number of months with excess precipitation were Himachal Pradesh, Haryana, Meghalaya, Punjab, Chandigarh and Andhra Pradesh in the five years preceding 1991, and Haryana, Jammu and Kashmir, Rajasthan, Himachal Pradesh and Punjab in the years preceding 2001.

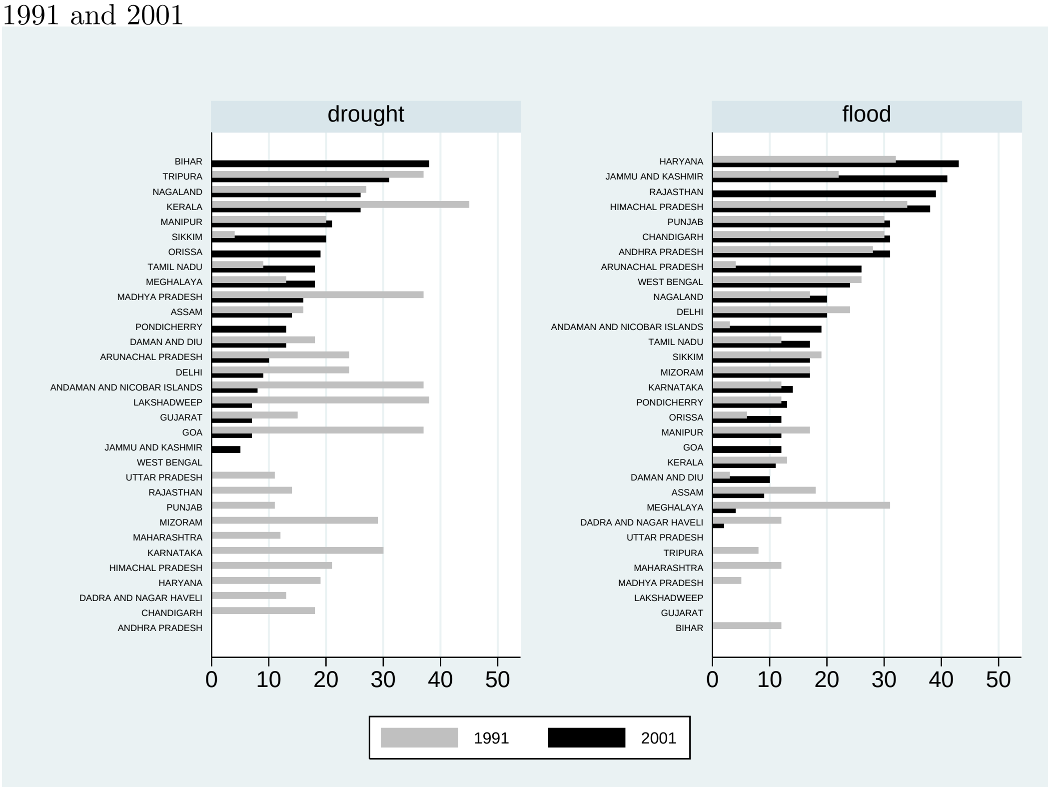
A comparison of the frequency of drought and excess precipitation with the migration data shows that the four states with the highest out-migration in the years studied (Uttar Pradesh, Bihar, Madhya Pradesh and Maharashtra) all experienced drought episodes, in particular the major out-migration states Bihar and Madhya Pradesh. By comparison, the states with high out-migration rates all had less than 12 months of moderate excess precipitation in the five years preceding the 1991 census and none in the five years preceding the 2001 census.

## A.6 Descriptive statistics

Table A.2 presents the mean, the standard deviation and the minimum and maximum of each variable. The total number of observations is 1860, representing bilateral migration between 31 Indian states in two years (1991 and 2001).

The average of 8 migrants per 100,000 individuals may seem very small, but the variable measures the bilateral rate for a unique origin-destination pair in one year. For example, 8 out of 100,000 individuals migrated from Assam to West Bengal between 1990 and 1991, which

Figure 3: Frequency of low precipitation ("drought") and excess precipitation ("flood") by state, 1991 and 2001




*Source:* Authors' own calculations based on the CRU TS3.21 data.
The definition of the frequency of low and excess precipitation is the number of months with the standardized precipitation index (SPI) at least one standard deviation below/above its long run mean.

represents a total of almost 1,800 individuals.[13] We have 930 possible combinations like this. It is also important to note that the dispersion is very large (the standard deviation is almost 4 times the mean) and that the bilateral migration rate can take values from 0 and up to 455 migrants per 100,000 individuals.

The average number of months (at any time) with drought or excess precipitation is almost 14 months (out of a total of 5*12 months), but the descriptive statistics show large variation in the variable, as indeed for all climatic factors tested here. The longest duration of a drought over the period studied was on average 12 months, just as for periods of excess precipitation. Over the time period studied the average drought and excess precipitation were of moderate size, but higher for droughts than for excess precipitation in the sum of the absolute values of the SPI (16.3 compared to 14.4).

[13]There are 22,408,756 individuals that did not move in 1990 from West Bengal.

Table A.2: Descriptive statistics and data sources

| Variable (n=1860) | Mean | Std. Dev. | Min. | Max. | Definition | Source |
|---|---|---|---|---|---|---|
| bilateral migration rate (x100,000) | 7.971 | 28.235 | 0 | 455.296 | $\frac{migration\_flow_{ij,t}}{population_{ii,t}}$ | Indian Census 1991 and 2001 |
| distance (km) | 1368.318 | 672.462 | 33 | 2846 | distance between states capital | www.mapsofindia.com |
| border (1/0) | 0.116 | 0.32 | 0 | 1 | dummy=1 if common border | www.mapsofindia.com |
| language (1/0) | 0.103 | 0.304 | 0 | 1 | dummy=1 if common language | www.mapsofindia.com |
| SC ratio | 1.006 | 0.111 | 0.776 | 1.288 | $\frac{SC\_rate_{jt}+1}{SC\_rate_{it}+1}$ | Indian Census 1991 and 2001 |
| ST ratio | 1.052 | 0.342 | 0.513 | 1.948 | $\frac{ST\_rate_{jt}+1}{ST\_rate_{it}+1}$ | Indian Census 1991 and 2001 |
| murder rate (x100,000) | 3.898 | 1.957 | 0 | 9.454 | $\frac{people\_murdered\_it}{total\_population}$ | www.systemicpeace.org/inscrdata.html |
| NSDP (USD) | 2.28e+09 | 3.18e+09 | 5598749 | 1.63e+10 | NSDP in 1980 constant prices | Reserve Bank of India |
| NSDP ratio | 24.023 | 88.472 | 0.001 | 1626.672 | $\frac{NSDP\_per\_capita_{jt}}{NSDP\_per\_capita_{it}}$ | Reserve Bank of India |
| irrigation rate (%) | 0.347 | 0.285 | 0 | 1 | $\frac{net\_irrigated\_area_{it}}{cultivated\_area_{it}}$ | Agricultural Census India |
| earthquake (1/0) | 0.016 | 0.126 | 0 | 1 | dummy=1 if earthquake(s) | EM-DAT |
| frequency of drought (#months) | 14.194 | 10.317 | 0 | 45 | #months with drought | CRU TS3.21 |
| maximum duration of drought (#months) | 11.855 | 9.949 | 0 | 37 | max. continuous # of months with drought | CRU TS3.21 |
| magnitude of drought (SPI) | 16.305 | 13.487 | 0 | 50.568 | sum of SPI during all months with drought | CRU TS3.21 |
| frequency of excess precipitation (#months) | 13.532 | 11.619 | 0 | 44 | #months with excess precipitation | CRU TS3.21 |
| maximum duration of excess precipitation (#months) | 12.048 | 10.266 | 0 | 43 | max. continuous # of months with excess precipitation | CRU TS3.21 |
| magnitude of excess precipitation (SPI) | 14.403 | 14.471 | 0 | 57.210 | sum of SPI during excess precipitation | CRU TS3.21 |
| temperature LR anomaly (°C) | 0.101 | 0.182 | -0.819 | 0.469 | z-score of temperature | CRU TS3.21 |
| precipitation LR anomaly (mm) | -0.001 | 0.056 | -0.141 | 0.148 | z-score of precipitation | CRU TS3.21 |
| flood frequency | 0.451 | 0.664 | 0 | 2 | #months with large flood events[a] | Dartmouth Flood Observatory |
| flood severity | 0.3991935 | 0.553 | 0 | 1.5 | class of flood[b] | Dartmouth Flood Observatory |
| flood magnitude | 0.964 | 1.381 | 0 | 3.793 | ln(frequency*severity) | Dartmouth Flood Observatory |

[a]Significant damage to structures or agriculture, long intervals since the last similar event, and/or fatalities.

[b]There are 3 classes. *Class 1, large flood events*: significant damage to structures or agriculture; fatalities; and/or 1-2 decades-long reported interval since the last similar event. *Class 1.5, very large events*: 20-100 years recurrence interval, and/or a local recurrence interval of at 10-20 years. *Class 2, extreme events*: with an estimated recurrence interval > 100 years.

# B Principal Component Analysis (PCA)

In order to match the climate and the census data, we have to aggregate the precipitation data to state level. The spatial grouping of observations is standard practice in the climatological literature (Munoz-Diaz and Rodrigo, 2004). These groupings serve to summarize climate data in a concise way (DeGaetano, 2001). PCA can be used to identify the most important correlations between different variables, so as to obtain a description of the major part of the overall variance, with a reduced number of linear combinations based on the original variables (Munoz-Diaz and Rodrigo, 2004). We apply PCA to test if aggregating precipitation across states implies losing important information or not.

We did a PCA between states and then between districts for the precipitation data, after having normalized the variables on the available period from 1901 to 2001. We applied an oblique rotation to the unrotated eigenvectors, according to the methodology of Barrios et al. (2010).[14] In the PCA applied to the states, we find 3 large rain zones with a loading of 0.1 (by having one single state which belongs to no zone and no state which belongs to more than one zone). By comparison, with a loading of 0.4, the precipitation patterns across states are completely independent, implying that there is no correlation between them (no regrouping of states were possible). The choice of the threshold for the loading is subjective: Singh et Singh (1996) take values included between 0.2 and 0.5, Barrios et al. (2010) take a value of 0.2 for their inter country analysis on sub-Saharan Africa and 0.05 for their intra country analysis; Munoz-Diaz and Rodrigo (2004) between 0.2 and 0.9.

We also check whether precipitation patterns are homogenous within states. When applying PCA to districts, we have 13 main rain zones with a loading of 0.1. The states contain between 1 to 3 different zones maximum, except for the states of Madhya Pradesh and Uttar Pradesh (regrouped in 5 and 6 zones respectively), but those are very large states. We checked the distribution of these zones on a map of India and confirmed that the states which belong to the same groups are indeed bordering, except in one case. We conclude that the analysis of climate variability at the state level seems relevant.

[14]Given that the PCA is for us only a preliminary analysis, we will not develop the technical details here, for further details see Munoz-Diaz and Rodrigo (2004), Barrios et al. (2010) and Singh and Singh (1996).

# C Additional tables

Table C.1: India migration classification. Census data (one-year duration).

| | 1990-1991 | | 2000-2001 | |
|---|---|---|---|---|
| | **Total** | **%** | **Total** | **%** |
| Intra-disctrict | 3,609,522 | 52% | 4,154,936 | 47% |
| Inter-district | 2,108,706 | 30% | 2,638,788 | 30% |
| Inter-state | 1,219,409 | 18% | 2,014,770 | 23% |
| **Total** | **6,937,637** | **100%** | **8,808,494** | **100%** |

Table C.2: Reasons for inter-state migration

| **Year** | **1991** | | | | **2001** | | | |
|---|---|---|---|---|---|---|---|---|
| **Female/Male** | **Female** | | **Male** | | **Female** | | **Male** | |
| Employment | 5272 | 10,85% | 22643 | 42,52% | 163951 | 18,79% | 620187 | 54,29% |
| Business | 1870 | 3,85% | 5954 | 11,18% | 13091 | 1,50% | 33665 | 2,95% |
| Education | 430 | 0,89% | 1100 | 2,07% | 430 | 0,05% | 1100 | 0,10% |
| Family moved | 19958 | 41,09% | 15570 | 29,24% | - | - | - | - |
| Moved after birth | - | - | - | - | 28439 | 3,26% | 31132 | 2,73% |
| Moved with household | - | - | - | - | 422965 | 48,48% | 282973 | 24,77% |
| Marriage | 9800 | 20,18% | 140 | 0,26% | 139115 | 15,94% | 2966 | 0,26% |
| Natural calamities | 200 | 0,41% | 200 | 0,38% | - | - | - | - |
| Others | 11044 | 22,74% | 7650 | 14,36% | 99234 | 11,37% | 146916 | 12,86% |

Table C.3: Correlation table

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) migration rate | 1.000 | | | | | | | | | | | | | | | |
| (3) distance | -0.278 | 1.000 | | | | | | | | | | | | | | |
| (4) border | 0.359 | -0.508 | 1.000 | | | | | | | | | | | | | |
| (5) language | 0.214 | -0.397 | 0.274 | 1.000 | | | | | | | | | | | | |
| (6) SC rate | 0.097 | 0.011 | -0.012 | -0.006 | 1.000 | | | | | | | | | | | |
| (7) ST rate | -0.101 | 0.039 | -0.027 | -0.009 | -0.670 | 1.000 | | | | | | | | | | |
| (8) murder rate | -0.067 | -0.042 | 0.008 | 0.090 | 0.127 | -0.196 | 1.000 | | | | | | | | | |
| (9) NSDP ratio | 0.138 | 0.020 | -0.053 | -0.042 | 0.321 | -0.256 | -0.151 | 1.000 | | | | | | | | |
| (10) irrigation rate | 0.097 | -0.079 | 0.003 | 0.016 | -0.438 | 0.315 | -0.007 | -0.191 | 1.000 | | | | | | | |
| (10) earthquake | -0.016 | -0.012 | 0.034 | -0.015 | 0.039 | 0.009 | -0.065 | -0.034 | -0.045 | 1.000 | | | | | | |
| *Drought variables* | | | | | | | | | | | | | | | | |
| (11) frequency | 0.044 | 0.090 | -0.031 | 0.011 | 0.086 | -0.029 | -0.017 | -0.032 | 0.013 | -0.151 | 1.000 | | | | | |
| (12) longest duration | -0.012 | 0.081 | -0.025 | 0.015 | 0.204 | -0.123 | 0.109 | 0.050 | -0.268 | -0.062 | 0.664 | 1.000 | | | | |
| (13) magnitude | 0.053 | 0.053 | -0.029 | 0.002 | 0.074 | -0.030 | -0.009 | -0.056 | 0.100 | -0.144 | 0.868 | 0.636 | 1.000 | | | |
| *Excess precipitation variables* | | | | | | | | | | | | | | | | |
| (14) frequency | 0.004 | 0.061 | -0.049 | -0.053 | 0.014 | -0.058 | 0.065 | 0.108 | 0.246 | -0.127 | -0.101 | -0.100 | 0.014 | 1.000 | | |
| (15) longest duration | 0.034 | -0.026 | -0.003 | 0.022 | -0.274 | 0.159 | -0.148 | -0.155 | 0.412 | -0.150 | -0.136 | -0.427 | -0.147 | 0.377 | 1.000 | |
| (16) magnitude | 0.015 | 0.046 | -0.049 | -0.034 | 0.048 | -0.096 | 0.084 | 0.104 | 0.166 | -0.114 | -0.048 | 0.013 | 0.136 | 0.923 | 0.290 | 1.000 |

Table C.4: Drought and excess precipitation with different lags

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Lag | 5 years | 4 years | 3 years | 2 years | 5 years | 4 years | 3 years | 2 years |
| ln distance$_{ij}$ | -0.677*** | -0.678*** | -0.677*** | -0.677*** | -0.679*** | -0.678*** | -0.678*** | -0.677*** |
| | (0.079) | (0.078) | (0.077) | (0.078) | (0.077) | (0.077) | (0.077) | (0.078) |
| border$_{ij}$ | 1.223*** | 1.219*** | 1.217*** | 1.217*** | 1.219*** | 1.215*** | 1.216*** | 1.217*** |
| | (0.149) | (0.148) | (0.148) | (0.148) | (0.147) | (0.147) | (0.147) | (0.148) |
| language$_{ij}$ | 0.400** | 0.397** | 0.399** | 0.398** | 0.400** | 0.400** | 0.400** | 0.400** |
| | (0.159) | (0.160) | (0.161) | (0.161) | (0.159) | (0.161) | (0.161) | (0.161) |
| $\ln \frac{SC_{jt}+1}{SC_{it}+1}$ | 1.590 | -0.032 | -1.390 | -1.682 | 13.544 | -5.295 | -5.043 | -1.044 |
| | (20.892) | (20.938) | (20.737) | (20.739) | (22.023) | (20.654) | (20.194) | (20.860) |
| $\ln \frac{ST_{jt}+1}{ST_{it}+1}$ | -2.498 | -2.520 | -2.868 | -3.037 | -4.227 | -3.778 | -3.860 | -2.903 |
| | (6.566) | (6.490) | (6.463) | (6.481) | (6.426) | (6.494) | (6.387) | (6.454) |
| ln murder pc$_{it}$ | 0.387 | 0.200 | 0.301 | 0.185 | 0.346 | 0.425 | 0.472* | 0.342 |
| | (0.287) | (0.304) | (0.298) | (0.290) | (0.287) | (0.277) | (0.268) | (0.286) |
| earthquake$_{it}$ | 0.636* | 0.552 | 0.563 | 0.421 | 0.532 | 0.584* | 0.619* | 0.586* |
| | (0.352) | (0.343) | (0.343) | (0.360) | (0.344) | (0.342) | (0.343) | (0.346) |
| drought frequency$_{it}$ | 0.016** | 0.009 | 0.002 | 0.017 | | | | |
| | (0.008) | (0.006) | (0.010) | (0.013) | | | | |
| flood frequency$_{it}$ | | | | | -0.014* | -0.008 | -0.012 | -0.004 |
| | | | | | (0.007) | (0.007) | (0.008) | (0.011) |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination-state/time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 |
| $R^2$ | 0.696 | 0.692 | 0.690 | 0.691 | 0.697 | 0.690 | 0.690 | 0.689 |

The dependent variable is the NSDP per capita ratio between state $j$ and state $i$ in year $t-1$.

Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

Table C.5: Income, drought and flood in India

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| drought frequency$_{it}$ | 0.004*** | | | | | |
| | (0.000) | | | | | |
| longest drought dur$_{it}$ | | 0.003*** | | | | |
| | | (0.000) | | | | |
| drought magnitude$_{it}$ | | | 0.002*** | | | |
| | | | (0.000) | | | |
| flood frequency$_{it}$ | | | | -0.004*** | | |
| | | | | (0.000) | | |
| longest flood dur$_{it}$ | | | | | -0.002*** | |
| | | | | | (0.000) | |
| flood magnitude$_{it}$ | | | | | | -0.002*** |
| | | | | | | (0.000) |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination-state/time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 |
| $R^2$ | 0.979 | 0.977 | 0.978 | 0.976 | 0.975 | 0.976 |

The dependent variable is the NSDP per capita ratio between state $j$ and state $i$ in year $t-1$. Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

Table C.6: Inter-state migration, drought and irrigation

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| ln distance$_{ij}$ | -0.676*** | -0.676*** | -0.676*** | -0.678*** | -0.676*** | -0.677*** |
| | (0.078) | (0.078) | (0.078) | (0.078) | (0.078) | (0.078) |
| border$_{ij}$ | 1.220*** | 1.217*** | 1.219*** | 1.220*** | 1.216*** | 1.218*** |
| | (0.148) | (0.148) | (0.148) | (0.148) | (0.148) | (0.148) |
| language$_{ij}$ | 0.402** | 0.401** | 0.400** | 0.403** | 0.401** | 0.405** |
| | (0.159) | (0.159) | (0.159) | (0.158) | (0.158) | (0.157) |
| ln $\frac{SC_{jt}+1}{SC_{it}+1}$ | -4.570 | -10.870 | -9.861 | -7.348 | -10.101 | -4.996 |
| | (20.562) | (21.046) | (21.102) | (19.843) | (21.152) | (21.011) |
| ln $\frac{ST_{jt}+1}{ST_{it}+1}$ | -2.656 | -3.397 | -2.823 | -1.251 | -3.293 | -1.349 |
| | (6.481) | (6.426) | (6.426) | (6.709) | (6.455) | (6.566) |
| ln murder pc$_{it}$ | 0.568* | 0.544* | 0.582* | 0.585* | 0.544* | 0.647** |
| | (0.312) | (0.316) | (0.318) | (0.311) | (0.315) | (0.323) |
| ln irrigation rate$_{it}$ | 2.059 | 2.472 | 2.402 | 1.382 | 2.432 | 1.550 |
| | (1.583) | (1.652) | (1.666) | (1.547) | (1.605) | (1.543) |
| earthquake$_{it}$ | 0.590* | 0.520 | 0.614* | 0.614* | 0.521 | 0.756** |
| | (0.352) | (0.344) | (0.354) | (0.355) | (0.344) | (0.368) |
| drought frequency$_{it}$ | 0.013* | | | 0.012* | | |
| | (0.007) | | | (0.007) | | |
| longest drought dur$_{it}$ | | 0.007 | | | 0.007 | |
| | | (0.006) | | | (0.006) | |
| drought magnitude$_{it}$ | | | 0.007 | | | 0.011** |
| | | | (0.005) | | | (0.005) |
| Interaction irrigation and drought | No | No | No | Yes | Yes | Yes |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination-state /time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 |
| $R^2$ | 0.698 | 0.696 | 0.696 | 0.699 | 0.696 | 0.699 |

Coefficients in the irrigation estimations are the net marginal effect of drought and irrigation. The dependent variable is the bilateral migration rate from state $i$ to state $j$ between year $t-1$ and year $t$. Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.

Table C.7: Inter-state migration and drought with bilateral fixed effects in an OLS model

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| $\ln \frac{SC_{jt}+1}{SC_{it}+1}$ | -21.872** | -24.247** | -22.721** | 0.555 | -4.649 | -1.577 |
| | (9.568) | (9.449) | (9.494) | (11.826) | (11.751) | (11.749) |
| $\ln \frac{ST_{jt}+1}{ST_{it}+1}$ | -9.733 | -10.261* | -9.844 | 0.744 | -0.529 | 0.529 |
| | (5.979) | (5.995) | (6.002) | (9.024) | (9.075) | (9.093) |
| ln murder $\text{pc}_{it}$ | 0.733* | 0.602 | 0.715* | 0.427 | 0.325 | 0.419 |
| | (0.384) | (0.393) | (0.384) | (0.377) | (0.381) | (0.377) |
| $\text{earthquake}_{it}$ | 0.681** | 0.575** | 0.682** | 0.634** | 0.522** | 0.649** |
| | (0.237) | (0.225) | (0.246) | (0.221) | (0.211) | (0.228) |
| drought $\text{frequency}_{it}$ | 0.017** | | | 0.018** | | |
| | (0.007) | | | (0.007) | | |
| longest drought $\text{dur}_{it}$ | | 0.013* | | | 0.013* | |
| | | (0.007) | | | (0.007) | |
| drought $\text{magnitude}_{it}$ | | | 0.008 | | | 0.009* |
| | | | (0.005) | | | (0.005) |
| Origin-state FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Destination-state FE | Yes | Yes | Yes | No | No | No |
| Destination-state/time FE | No | No | No | Yes | Yes | Yes |
| Bilateral FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | No | No | No |
| N | 1860 | 1860 | 1860 | 1860 | 1860 | 1860 |
| $R^2$ | 0.809 | 0.809 | 0.809 | 0.828 | 0.828 | 0.828 |

The dependent variable is ln(bilateral migration rate+1) from state $i$ to state $j$ between year $t-1$ and year $t$. Robust standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.001$.