WATER GLOBAL PRACTICE

Climatic Shocks and Internal Migration

Evidence from 442 Million Personal Records in 64 Countries

Guy J. Abel, Raya Muttarak, and Fabian Stephany

WORLD BANK GROUP

GWSP
GLOBAL WATER SECURITY & SANITATION PARTNERSHIP

About the Water Global Practice

Launched in 2014, the World Bank Group's Water Global Practice brings together financing, knowledge, and implementation in one platform. By combining the Bank's global knowledge with country investments, this model generates more firepower for transformational solutions to help countries grow sustainably.

Please visit us at www.worldbank.org/water or follow us on Twitter at @WorldBankWater.

About GWSP

This publication received the support of the Global Water Security & Sanitation Partnership (GWSP). GWSP is a multidonor trust fund administered by the World Bank's Water Global Practice and supported by the Australian Department of Foreign Affairs and Trade, Austria's Federal Ministry of Finance, the Bill & Melinda Gates Foundation, Denmark's Ministry of Foreign Affairs, the Netherlands' Ministry of Foreign Affairs, the Swedish International Development Cooperation Agency, Switzerland's State Secretariat for Economic Affairs, the Swiss Agency for Development and Cooperation, and the U.S. Agency for International Development.

Please visit us at www.worldbank.org/gwsp or follow us on Twitter at @TheGwsp.

Climatic Shocks and Internal Migration

Evidence from 442 Million Personal Records in 64 Countries

Guy J Abel,[1,2] Raya Muttarak,[2] and Fabian Stephany[3]

[1]Asian Demographic Research Institute, Shanghai University, China

[2]International Institute for Applied Systems Analysis, Austria

[3]Oxford Internet Institute, University of Oxford, United Kingdom

WORLD BANK GROUP

Climatic Shocks and Internal Migration: Evidence from 442 Million Personal Records in 64 Countries

Abstract

This paper examines whether and how climatic shocks influence individual migration decisions. We use census microdata across 64 countries over the period 1960 to 2012, covering 442 million individual records, combined with geo-referenced temperature and precipitation data summarised for each origin and destination administrative unit. Migration is identified when an individual changed a place of usual residence one, five or ten years ago to a new major administrative unit in the same country. Given an exceptionally large number of observations, we apply a two-step approach to analyse the relationship between exposure to climatic shocks and migration. First, we use random forest models to uncover that in many countries climatic shocks are as important as better-known individual-level covariates in determining migration decisions. This observation serves as a yardstick for the second step of our analysis. For a subset of countries, where rainfall shocks play an important role in migration, we compare internal migration patterns across time by examining whether a region experiencing positive or negative rainfall shocks observed higher or lower migration. We find that negative rainfall shocks suppress outmigration particularly for low-income countries. The opposite is true for positive rainfall shocks whereby migration is found to increase, especially for lower-income countries. Our finding supports the liquidity constraint argument whereby adverse climatic conditions can disrupt migration financing and consequently suppress ability to migrate.

Keywords: census micro data, climatic shocks, drivers of migration, internal migration, machine learning, random forest models

Introduction

The past decade has observed a significant increase in scientific studies on the relationship between climatic shocks, natural hazards and migration (Borderon et al. 2019; Hoffmann et al. 2020; Piguet et al. 2018). Earlier studies in the mid-1980s to 2000s assume an 'alarmist' approach which view the environment as a principal driver of population movements and often predict a large-scale migration of hundreds of millions of affected populations (El-Hinnawi 1985; Myers 1993, 2002). The forecasts for the number of climate migrants or refugees by 2050 range from 25 million to 1 billion (IOM 2014). However, the figures given are not grounded in empirical realities but rather are speculative common sense (Gemenne 2011). Despite the great publicity such work gained among the media and policy-making community, the scientific community is sceptical about the approach and calls for more sound data and methods to measure and project environmental migration (Piguet 2010). It is emphasised that migration

Acknowledgement: This is a background paper for the report titled *Ebb and Flow, Volume 1: Water, Migration, and Development*. We would like to sincerely thank Esha Dilip Zaveri, Jason Daniel Russ, and Amjad Muhammad Khan from the World Bank for their intellectual inputs and support throughout the preparation of this manuscript.

is a complex process. Black et al. (2011) highlight that environmental and climatic conditions are rarely a sole and direct driver of migration but they may affect migration through altering other social, economic, political and demographic factors underlying migration decision.

Recent empirical studies of environmental migration have made progress in addressing the complexity of migration by using statistical methods to examine the interactions between various micro, meso and macro drivers of migration while attempting to isolate environmental effects on migration (Fussell et al. 2014). A growing body of empirical research has challenged the narrative of climate-driven mass migration as portrayed by the 'alarmist' literature. Given different contexts and type of migration studied, the evidence on the direction and the extent to which environmental and climatic factors influence migration is mixed: adverse climatic conditions induce migration responses in some cases but suppress migration for some others (Borderon et al. 2019; Cattaneo et al. 2019; Hoffmann et al. 2020). Heterogeneity in migration responses to climatic shocks is shaped not only by the types of climatic events (e.g. slow vs. fast onset) but also underlying factors determining vulnerability, adaptive capacity and ability to migrate including wealth, gender, age, health and financial and human capital (Cattaneo et al. 2019).

Despite inconsistencies in the findings, one empirical regularity is that climate-induced migration is likely to be a short distance move within national borders rather than international migration. This is in line with migration as a whole, where the vast majority of migration occurs internally rather than across international borders; Bell and Charles-Edwards (2013) for example, estimated approximately 763 million persons living within their own country but outside their region of birth in 2005, compared to 214 million international migrants at the same time. Accordingly, recent studies have shifted from the narrative of mass migration of climate migrants crossing borders to enter to Europe and North America to focus on internal migration (Rigaud et al. 2018). Most existing empirical studies on climate-related internal migration focus only on a single country and consequently yield substantial heterogeneity in the findings across studies (Dillon et al. 2011; Gray and Mueller 2012a, 2012b; Mastrorillo et al. 2016; Mueller et al. 2014). This is partly due to differences in climate and migration measures used, leading to comparative internal migration studies across country contexts using consistent data, measurement and methods (Gray and Wise 2016; Mueller, Gray, et al. 2020; Mueller, Sheriff, et al. 2020; Thiede et al. 2016). However, in these examples only a few countries are included, and the studies are limited to a specific world region. Larger-scale cross-national studies of the climate effects on internal migration (Barrios et al. 2006; Beine and Parsons 2015; Cattaneo and Peri 2016; Henderson et al. 2017; Marchiori et al. 2012) have been carried out, but use urbanisation as a proxy for migration, thus only rural-urban moves are captured and other types of moves such as rural-rural and urban-urban moves are missing from the analyses.

Exploiting census data on migrants from the Integrated Public Use Microdata Series International (IPUMSI) database (Minnesota Population Center 2019) and gridded climate data (Willmott and Matsuura 2015), we analyse the association of exposure to climatic shocks and internal migration for 64 countries. To the best of our knowledge, ours is the first study to use census microdata for a large body of countries to analyse climate-related migration, taking up the call for more comparative migration research, over time and across population subgroups as suggested by Sobek (2016). The large volume of data obtained from the harmonised census microdata, with 442 million individual records in our case, requires a new approach to handle and analyse the data, where classical regression methods typically used for the study of climate and migration are unable to handle so many observations.

To this end, we employ a two-step approach. In the first step, we use a machine learning filter with a random forest model for the identification of relevant populations that use migration as a response to exposure to climatic shocks. This non-parametric technique allows us to identify patterns in the data without imposing any statistical assumptions whilst being sensitive to misuse of significance testing that might occur with conventional multiple regression models (Attewell et al. 2015). Upon identifying climatic factors as an important driver of migration, in the second step of the analysis, we conduct a comparison of regional migration shares across time in order to identify the direction of the climate effects on migration.

Our study has multiple contributions to the empirical knowledge in the field of climate change and migration. First, we consistently analyse the effects of climatic shocks on internal migration – the most common form of climate-related migration – across many countries. We overcome this issue by using harmonised census microdata for 64 countries covering major continents including Africa, Asia, Europe, North America and South America. Second, based on individual-level data, our study controls for relevant individual characteristics determining migration while examining the importance of climate drivers. This allows us to overcome the 'ecological fallacy' of looking only at macro-level correlations of push-pull factors without linking with individual migration responses, in line with De Haas (2011, p. 16) observation that 'people do not migrate 'because of' abstract concepts such as demographic transitions, declining fertility, ageing, population density, environmental degradation or factor productivity.' Third, using a novel machine learning technique, we analyse large-scale individual records of 442 million individuals and identify the important drivers of migration without imposing any statistical or theoretical assumptions. We overcome the constraint of the random forest models, namely their limitation in specifying the directions of the effect of the individual predictors on the outcome variable identified in a previous study of environmental migration (Best et al. 2020), by analysing a subset of regions in the sample. Selecting a pair of subnational regions which appear in the data in at least two time periods: one with exposure to climatic shocks and one without, we are able to compare changes in migration levels in each region. This additional step allows us to focus on regions where climatic factors are identified as an important driver of migration and single out the direction to which climatic shocks positively or negatively influence migration patterns.

The remainder of the paper is organised as follows. The next section describes the individual-level microdata and climate data employed for the analysis. Section 3 elaborates on the method of random forest models in our two-stage filtering technique. Section 4 describes the climate effects on internal migration patterns and the heterogeneity by climatic factors and country contexts. The final section summarises the key findings and concludes on how our findings and methods could be expanded in the future.

Data

Individual migration and socio-demographic characteristics data

Individual data on migration and socio demographic characteristics were obtained from harmonised census microdata samples from the Integrated Public Use Microdata Series International (IPUMSI) database (Minnesota Population Center, 2015). IPUMSI provides the world's largest archive of publicly available census samples, with variables harmonised across countries and over time to facilitate comparative research. Samples in IPUMSI are typically close to 10 percent of the entire census; see Table A1 in the appendix for the sampling fractions of the IPUMSI samples used in our study.

Over 280 census samples in IPUMSI contained variables related to migration. We selected only the samples with variables associated with responses to questions on the individual's previous region of residence. The identification of migrants, given the availability of a response to a question in individuals' past region of residence was derived in either one of the two following approaches. In the first approach we used an IPUMSI variable on an individual's region of residence at a fixed period prior to the census, such as one or five years ago. If the region was different from the region of the respondent's household (provided in another IPUMSI variable on the region of respondent at the time of the census) then we coded the respondent as a migrant. If a variable on a fixed period prior to the census was unavailable, we used a combination of three variables; the first variable contains the respondent's previous region of residence, the second variable indicates if the previous residence of the respondent was in the same or different region to the current location and the third variable gives the number of years the respondent has resided in their current location. If the respondent had changed their location within a year, and the previous residence was different to the current location (based on the second and third variables), then we coded the respondent as a migrant. We used the first variable on the previous region of residence to identify the climatic conditions for the migrant origin region, as discussed later. For the IPUMSI census samples where we derived migrants using the latter approach are indicated using the 1* in Table A1, to indicate an implied period of migration within a year from the respective census.

Under both approaches to deriving migrants we are identify migration events based on transitions rather than the movements. In other words, we derived migrant events from a comparison of the regions of residence at the start and end of a given interval for each individual. We do not capture multiple changes of residence that might have occurred within the period or region, and so, for example we would miss a migrant who moved to a different region and then returned to their original region within the migration interval. At a population level, the difference between the number of migration events and migration transitions will be small when the interval length is short, for example one-year. The discrepancy increases in a non-linear fashion as the interval length increases; see for example Rees (1977) or Courgeau (1973) for further details on the differences between migration transitions and movements.

Measurement of migration

We were able to create a binary migrant variable for individuals in 189 IPUMSI census samples listed in Table A1. Although there were more IPUMSI census samples with migrant variables, an additional variable on the origin region, that we used for identifying climate conditions of migrants in our analysis were not available in other samples. In total our selection of IPUMSI census samples cover 441.6 million individual records, from which 18.7 million were coded as migrants. Summary details on the number of records and migrants by country and census year can be found in Table A1.

Direct cross-national comparisons between numbers of migrants, migration rates or the probability of migration are not possible due to the different census sample sizes and the different set of regional geography in each country, where countries are subdivided into different numbers of regions and regions themselves can vary greatly in both geographic and demographic scales. Consequently, the number of migrants is likely to be relatively higher in countries with more regions and large population in comparison to similar country with fewer regions and a smaller population (Bell, Charles-Edwards, Kupiszewska, et al. 2015; Bell, Charles-Edwards, Ueffing, et al. 2015) - a feature related to the Modifiable Areal Unit Problem (MAUP). In all countries we used a harmonised first level administrative regional geography provided by IPUMSI.

Measurement of socio-demographic characteristics

In addition to deriving migrant status for all individuals in the census samples, we also used five harmonised IPUMSI variables on individual characteristics available in all 189 census samples: age, sex, education, marital status and the number of persons living in the household. These are the variables relevant to the migration decision identified in previous literature.

The evidence on an **age** profile of migration is relatively consistent across countries. Migration rates by age are generally homogeneous across different populations, with higher rate of migration in early adulthood (e.g. for education, taking up an employment and starting up a family) and lower rate of migration upon exit from the labour market (Rogers and Castro 1981) although variations are known to exist (Bernard et al. 2014; Bernard and Bell 2015). Migration decision commonly varies along the life-cycle model where life-course transitions such as leaving parental home, marriage and labour market participation determine patterns of movement (Aybek et al. 2015). The role of age on environmental migration is consistent with general migration patterns with the young and middle-aged having higher intention and higher propensity to migrate (Borderon et al. 2019).

There are however scarce empirical regularities with respect to migration responses by **gender** (Cattaneo et al. 2019; Chindarkar 2012). Migration response is linked with gender disparities in vulnerability to and ability to respond to and cope with climate change. On the one hand, if women are more vulnerable to climate change than men, the stronger impact of climate change on their livelihoods can trigger migration as an adaptation response. On the other hand, women may have lower ability to use migration as an adaptation option, especially if they have lower opportunities in the labour market compared to men. Accordingly, there is no clear gendered pattern on migration response to environmental and climatic shocks.

Education is a key determinant of both international and internal migration (Ginsburg et al. 2016). However, there is no conclusive evidence on whether migrants are drawn upon a pool of less or more educated individuals (N. Williams 2009). Empirical studies show that the direction of migrant selectivity by educational level differs across countries (Cattaneo 2007; Gould 1982). Previous studies at the individual level provide inconsistent evidence on the relationship between education and the propensity to migrate. On the one hand, a series of studies report a positive effect of educational attainment on the likelihood of internal migration both for males (Agesa and Agesa 1999; Chae and Glick 2019) and females (Brockerhoff and Eu 1993; Chae and Glick 2019). On the other hand, many studies find a negative relationship between education and migration (Massey et al. 1987; Massey and Espinosa 1997; Quinn and Rubb 2005; Reed et al. 2010) and some studies reported no significant association at all (Adams 1993; Curran and Rivero-Fuentes 2003). Whether highly educated individuals are more or less likely to migrate following climatic shocks also depends on their adaptive capacity and adaptation options available. Whilst human capital enables *in situ* adaptation, it also facilitates migration. Education differentials in vulnerability and adaptive capacity may explain the heterogeneity in the effect of education on migration.

Marital status is also an important individual characteristic determining migration. Typically, unmarried individuals are more mobile than married persons. The latter are less likely to migrate on their own and have higher chance of returning to the place of origin sooner (Simpson 2017). It has been shown that the likelihood of individual migration declines following marriage (Kanaiaupuni 2000; Kuhn 2005). In the context of climate change, one may expect women being more likely to migrate for marriage reason. There is evidence that parents use daughter's marriage as a risk minimisation strategy (Rosenzweig and Stark 1989): for instance, to reduce household size and food consumption in times of crisis as evident in drought-induced migration in rural Mali (Findley 1994). If this is the case, then single women are more likely to migrate as a response to climatic shocks.

Household characteristics, including **household size** and structure, are an important determinant of migration. Not only do larger households have more options to diversify income sources and minimise risks (de Haan 1997; Tsegai 2007), but they also have higher pressure to deal with scarce resources in times of economic or environmental shocks. The number of people in the household is associated with more options for the arrangement of the labour resources. Therefore, migration probability is likely to increase with household size since larger households have greater surplus labour (Phuong et al. 2008).

Age and the number of persons living in the household (**household size**) are continuous variables. **Sex** is a binary variable classified into male and female. The harmonised **education** variable contains four categories: less than primary, primary completed, secondary completed and university completed. The harmonised **marital status** variable also contains four categories: single/never married, married/in union, separated/divorced/spouse absent and widowed.

Climate data

Climatic shocks are measured as rainfall deviations from the long-term panel mean of rainfall or temperature for a given country. The deviations are measured as standardised Z-scores which allow us to measure climatic differences in terms of the historical range of climatic variability across heterogeneous areas. Z-scores tend to be better predictors of migration outcomes compared to raw climate data

(Gray and Wise 2016; Mueller, Sheriff, et al. 2020). We calculate subnational-level annual rainfall Z-scores for each census-year. Positive Z-scores indicate wetter than normal conditions while negative values indicate drier than normal conditions.

Rainfall data are obtained from the gridded monthly time series data from Willmott and Matsurra (University of Delaware) V 4.01 for the years 1900-2014 (Willmott and Matsuura 2015). The time-series data for rainfall consist of monthly average precipitation spanning from the period 1900 to 2014 calculated on high-resolution (0.5 x 0.5 degree) grids. The precipitation grids are used to calculate rainfall Z-scores for each administrative unit matching the place of origin indicated by the previous place of residence of the migrants.

Depending on the migration interval defined in each census (place of residence 1, 5 or 10 years preceding the census), corresponding climatic conditions were calculated based on the length of the migration period and an additional two years before, i.e., 3, 7 and 12 years prior to the census respectively. For instance, for the Iraq Population and Housing Census 1997, the participants were asked about their place of residence 10 years ago. Accordingly, the rainfall Z-scores for the entire population (migrants and non-migrants) are calculated based on average rainfall of the period 1984-1996 and the long run average based on the complete historical series.

Methods

The combined micro censes and climate data cover with 442 million individual records from 1,329 regions across 189 censuses covering 64 countries. This exceptionally large volume of data makes it inappropriate to estimate migration drivers using classical regression methods.[2] Consequently, we employ a two-step approach. First, we make use of random forest models on a country level to identify populations that were particularly sensitive to rainfall shocks in their migration behaviour. This non-parametric technique allows us to identify patterns in the data without imposing any statistical assumptions and being sensitive to misuse of significance testing like in conventional multiple regression models (Attewell et al. 2015). In a second step, we focus our analysis to the climate-sensitive countries, for which climate shocks were as important as individual characteristics and compare changes in internal migration shares in a given region across time. Region pairs are identified when there are two occurrences of the same region in the dataset but at different points in time (i.e., different censuses): one with and one without a rainfall shock. The share of internal migrants for each region and year is calculated and the two years are contrasted for each region by calculating the difference between the migration share in a year with and in a year without an extreme event. A positive value indicates that more people had migrated in a year when a shock had occurred.

Random Forest Analysis

In the first step of our analysis, a random forest model is performed for each of the 64 countries in our dataset. Random forest is a robust machine learning algorithm that can be used for a variety of analyses

2 It is argued that measures of statistical significance such as p-values, lose meaning for large sample sizes well above 10,000 observations (Lin et al. 2013).

including regression and classification. The technique is suitable for our context with a large dataset where we aim to identify important covariates underlying migration decisions. It is an ensemble method, meaning that a random forest model is made up of many small decision trees, which each produce their own predictions (Hastie 2009).

The individual decision trees apply a systematic truncation of a data sample based on the distribution of an outcome variable (the target feature). The target feature in our case is internal migration, which distinguishes each individual observation between migrants (labelled 1) and non-migrants (labelled 0). The target feature in an initial sample of each country's population has a given distribution, e.g., 10 percent of a country's population migrated and 90 did not migrate. For this initial distribution, a measure of data purity, in our case the Gini impurity, is calculated. The Gini impurity provides a measure of how often a randomly chosen element from the sample would be incorrectly labelled if it were randomly labelled according to the distribution of labels in the sample. This measure serves as a benchmark for the subsequent steps.

In each decision tree, our features, personal characteristics and regional climatic shocks, are considered for the truncation of the initial sample. The data purity of the two resulting subsets (nodes) are compared. The external feature that yields the highest level of data purity in the remaining subsets are chosen as a splitting criterion. After the split, the initial procedure is repeated for each of the resulting subsamples, testing all features: a tree-like, cascading structure emerges after repeating the procedure several times.

More formally, at each node, τ, within the decision tree the optimal split is sought using the Gini impurity $G(\tau)$. If we have C possible outcomes of our target feature, in our case 0 (for no migration) and 1 (for migration), and $p(i)$ is the probability of choosing a datapoint with class i, then the Gini impurity is calculated as:

$$G(\tau) = \sum_{i=1}^{C} p(i) * (1 - p(i))$$

Formally, we can measure the change in our Gini impurity measure $\Delta G(\tau)$ after a truncation by one of our input features, i.e., splitting the sample into subpopulations below and above a certain age or education level.

In a random forest model, randomly selected decision trees are combined (Hastie 2009). Each node in the decision tree works on a random subset of features to calculate the output. The random forest then combines the output of individual decision trees to generate the final output. In this ensemble of decision trees an exhaustive search over all features θ available at the node the maximal $\Delta \mathrm{G}(\tau)$ is determined. The decrease in Gini impurity resulting from this optimal split $\Delta G_\theta(\tau, T)$ is recorded and accumulated for all nodes τ in all trees T in the forest, individually for all variables θ:

$$MDG\ (\theta) = \sum_{T} \sum_{\theta} \Delta G_\theta(\tau, T)$$

This quantity, the Mean Decrease Gini (MDG), calculates by how much the performance (Gini) of the random forest model would deteriorate, on average, if a respective feature were to be excluded from the model (Hastie 2009). A high MDG indicates that the respective characteristic is important when truncating the sample regarding the target feature, like a change in R^2 in a regression model.

Our random forest model is fed with a ten percent sample of each country's population and the accuracy of the random forest with 1,000 tree combinations is evaluated on a 25 percent test set, a typical split for cross validation. With MDGs for all features, individual and climate characteristics, the relative MDG (RMDG) can be calculated for each feature θ.

$$RMDG_{\theta,\eta} = MDG_{\theta}/MDG_{\eta} * 100$$

This value of the RMDG assesses the importance of a given feature, θ, relative to the importance of another reference feature. For our purpose, the MDG of the climatic shock characteristic is compared with the MDG of the education variable, as education is known to be a relevant factor for migration behaviour (Ginsburg et al. 2016). A RMDG of more than 100 indicates that a given climatic shock is more relevant for explaining migration patterns than educational attainment. As a selection threshold, only cases with a climatic shock MDG of higher than 50, e.g., climatic shocks being at least half as important as education, are marked as sensitive populations and considered in the following regional pairwise comparison.

Why machine learning

Multiple regression analysis is a very popular techniques used in policy-related research. Regression models wide-spread use is in part due to their ability to quantifying relationships between variables of interest in a custom and interpretable way, e.g., estimated coefficients of explanatory variables allow for making statements about the magnitude (size of the coefficient) and relevance (significance level) of the relationship. However, a very large sample size impedes an application of regression models because the expressiveness of significance levels diminishes for samples larger than 10,000 observations (Lin et al. 2013). Classification techniques in machine learning techniques, such as random forests, on the other hand, do not rely on variances and are applicable also on larger samples. Among machine learning classification techniques, random forests come with a data purity metric i.e., the mean decrease in Gini that allows us to compare the influence of different features for the classification, akin to changes in R squares in a multivariate regression analysis.

Pairwise Regional Comparison

Having identified countries with a climate sensitive population via the random forest models, we perform a pairwise regional analysis of annual shares of migrants. For relating climatic shocks and migration on a regional level, we calculate differences in migration shares. First, region pairs are identified, i.e., the same region in the dataset but at different points in time, one with and one without a climatic shock. In a second step, the share of internal migrants for each region and period is calculated and then the two periods are contrasted for each region by calculating the difference between the migration share in the period with and without an extreme event. A positive value indicates that more people had migrated in a year when a shock had occurred.

Formally, the change in migration shares β for region i can be calculated by comparing migration shares in periods with a climatic shock $t(\gamma = 1)$ to shares in a period without a shock $t(\gamma = 0)$ event:

$$\Delta\beta_i = \beta_{i,t(\gamma = 1)} - \beta_{i,t(\gamma = 0)}$$

The advantage of this approach is that regions can be related individually by the potential impact of a climatic shock and their contextual characteristics, such as the level of aridity or GDP per capita. Likewise, the potential direction of the relationship between migration and climatic shocks can be assessed on a regional level. Similarly, it relates every regional change to the economic and environmental characteristic in that given point in time. As a disadvantage, the model does not allow us to control for individual features that might influence migration

The presented two-step approach, a reduction of the sample population via random forest analysis followed by a regional pairwise comparison, is helpful when large sample sizes impose statistical hurdles on established inferential models. In addition, the analysis presented here also allows a closer examination of the causal relationship between migration and extreme climate events while it is not possible to do so using multivariate regression analysis or random forest models alone. The machine learning-driven pre-selection of countries does not make any claims on causal effects, but it allows to effectively filter out a set of regions where climatic factors could play a causal role for migration. The pairwise comparison measures changes in migration for the same region over time. This brings analysis close to a “twin-like” experimental setting and limits the influence of potential regional confounders that could have influenced migration patterns. A future refinement of this process could consider regional characteristics that might have changed over time and therefore could have impacted migration in addition to climatic shocks.

Results

Individual and environmental drivers of migration

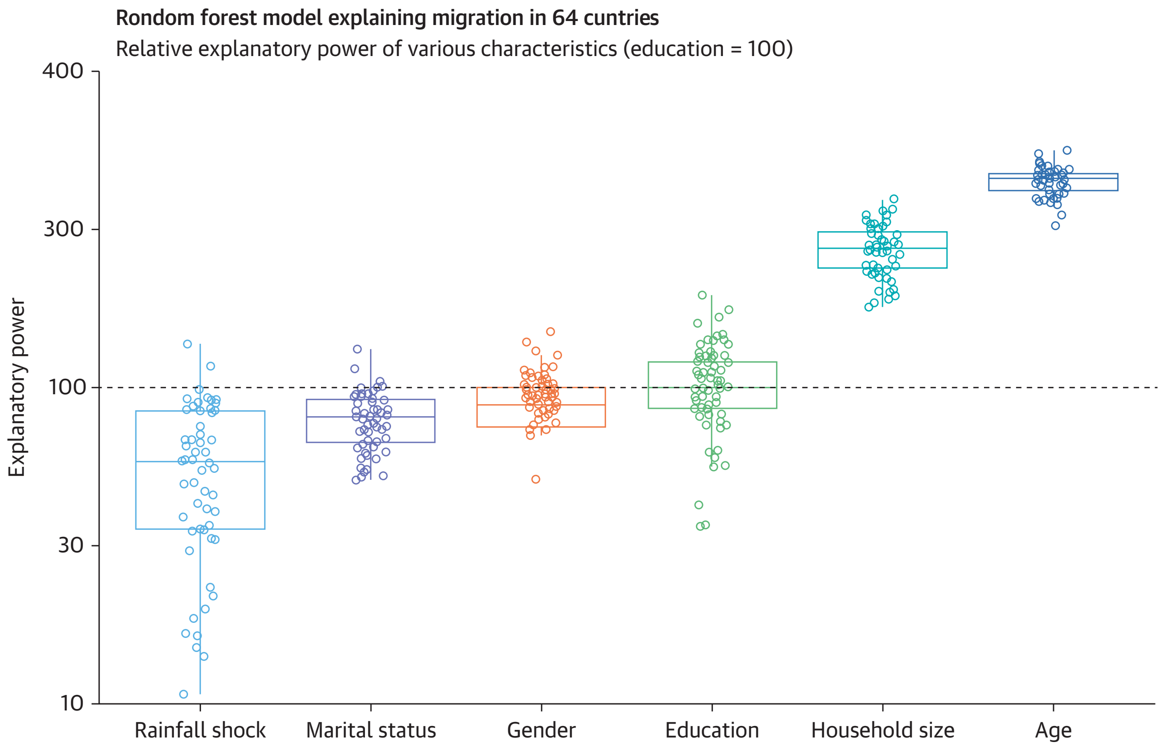
As part of our first analysis round, we apply a random forest model that allows us to compare the importance of each of the individual characteristics and exposure to climatic shocks for migration for 64 countries in our dataset.

Figure 1 summarises the results of our random forest model for all six characteristics (five individual features and a climatic shock occurrence) across 64 countries.[3] Each dot shows how important a respective characteristic is for explaining migration in each country. The ‘explanatory power’ (RMDG) of each characteristic in a given country is shown relative to the importance that education has, on average, across all countries (100 = dotted line). We selected education as the comparison variable, due to the ample evidence of its factor in migration decisions (Dustmann and Glitz 2011),

There are clear differences in explanatory power across characteristics. While age and household size are, on average, more important than education, other characteristics like sex or marital status are,

3 This analysis has been carried out six times in total, each for one of the three climatic shock characteristics with two different magnitudes. Figures A1 and A2 show the results for positive temperature and precipitation shocks.

FIGURE 1: Important drivers of migration identified by random forest models for 64 countries



on average, as relevant as education. The climatic shock analysed in this model, i.e., the occurrence of a negative precipitation shock, is slightly less important than education (51% of education's importance). However, the widespread distribution of the climate characteristic indicates that negative precipitation shocks are as important as sex, marital status or education, in some countries.

For each country, the importance of each of the six features, i.e., five individual characteristics and one contextual variable i.e., rainfall shock, when explaining individual migration can be compared. A ranking of the importance of each feature shows that age and household size usually have the highest MDG, followed by education, gender, and marital status. The rainfall shock feature is usually ranked as the least relevant of all features. This is of little surprise, as the five individual characteristics are known to be relevant drivers of migration. However, in contrast to individual characteristics, climate shocks vary strongly in their importance in explaining migration. For a set of countries shown in Figure 1[4], precipitation shocks have a higher feature importance than some individual characteristics. For some of these countries, mostly less economically affluent countries, the rainfall shock was ranked as the fourth (Colombia, Jamaica) or third (Botswana, Canada, Trinidad and Tobago) most important explanatory factor of all features.

4 These countries are Botswana, Canada, Trinidad and Tobago, Columbia, Jamaica, Bulgaria, Costa Rica, Cuba, Egypt, Spain, Guinea, Kenia, Malaysia, Nepal, South Sudan, Tanzania, and Vietnam.

Pairwise Regional Analysis

In the subsequent analysis, only the countries with explanatory power values above 50, e.g., half of the average importance of education are included in the pairwise regional comparison. As our dataset spans across several census years, we can observe many regions at different points in time. Some of these regions experienced a climatic shock in one census, while not witnessing extreme climate conditions in another census. This allows us to compare the same region with and without the occurrence of a climatic shock[5]. For these 'region pairs' we calculate the change in migration (regional share of migrants in a census with a shock - regional share of migrants in a census without a shock). A positive value, for example, indicates that a larger share of the region's population had migrated in the census where climatic shocks occurred. We only consider region pairs for which this change in migration was sufficiently large i.e., where the changes in migration are statistically significant at a 95% confidence level in each GDP quartile.

Figure 2 groups regions by their level of income as measured by log GDP per capita. Blue dots represent regions for which a larger share of people migrated in years of climatic shocks, while red dots stand for regions in which a smaller fraction of the population had migrated during a shock year. Table A2 provides a summary of the distribution of the paired regions by income level of the country and type of precipitation shocks.

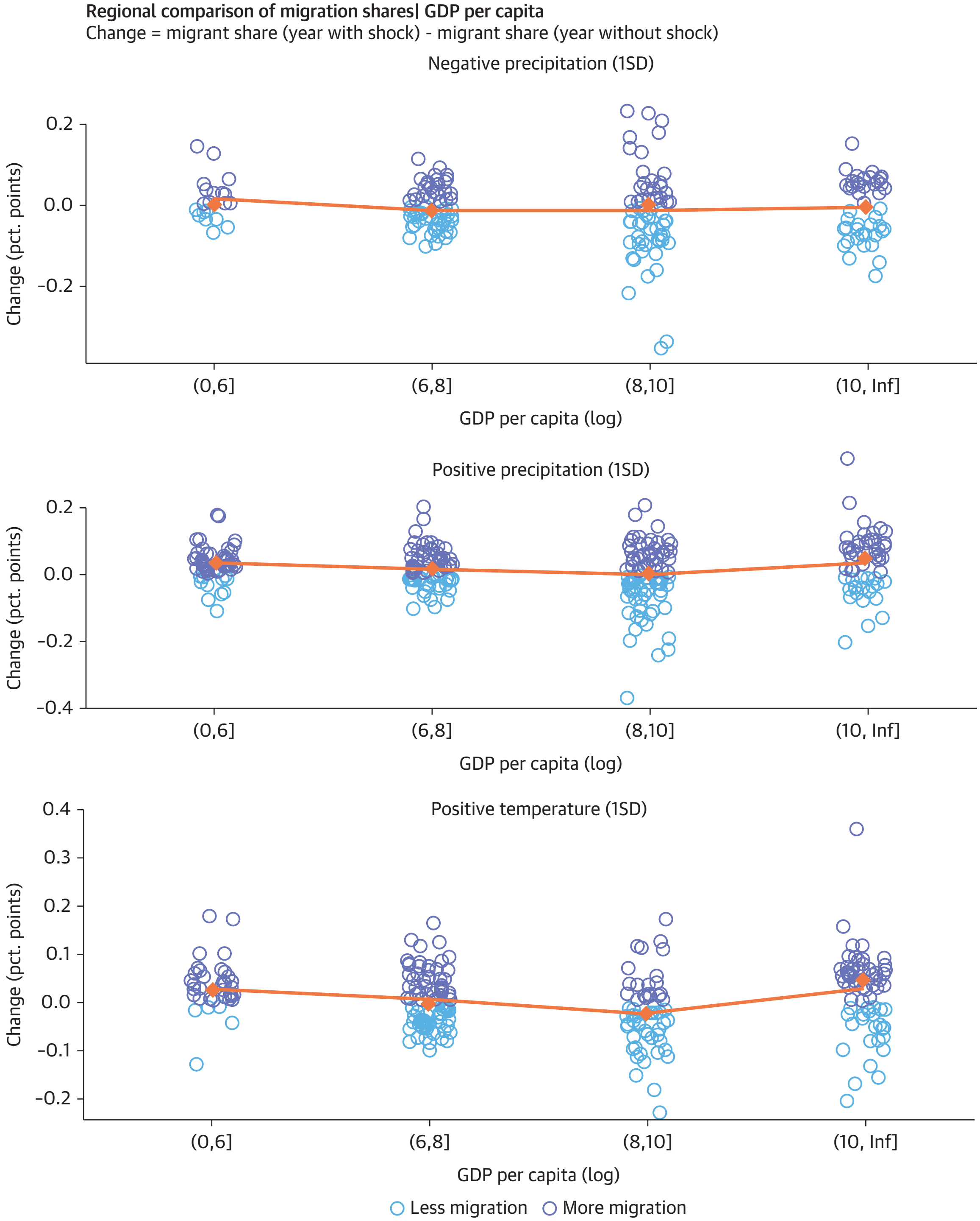
It is evident that for all income levels, positive precipitation shocks are associated with increase in migration, but the relationships appear to be particularly evident in low-income countries. Of the 136 regions in the lowest income group, 136 (73%) experienced more migration in periods with a positive precipitation shock (1SD). This is also the case for the high-income group where over 60% of the regions experienced an increase in migration after exposure to a positive precipitation shock. In contrast, migration declines when a region experienced negative precipitation shocks and this pattern is particularly evident amongst regions (60%) located in low-income countries.

TABLE 1: Regional comparison of changes in migration shares by precipitation shocks and GDP per capita

Precipitation shock	GDP per capita	Less migration		More migration		Total
		N	%	N	%	
Negative precipitation (1SD)	Low	80	60.2%	53	39.8%	133
	Middle	34	47.2%	38	52.8%	72
	High	29	50.9%	28	49.1%	57
Positive precipitation (1SD)	Low	51	27.3%	136	72.7%	187
	Middle	54	48.2%	58	51.8%	112
	High	21	35.6%	38	64.4%	59

5 The observed patterns are similar for more extreme climatic shocks, e.g., two standard deviations, but weaker as less region pairs have been identified. The results are shown in Table A1 in the Appendix.

FIGURE 2: Regional comparison of changes in migration shares on GDP per capita log scale



Discussion and conclusion

Applying machine-learning techniques to analyse drivers of migration, this is the first study to comprehensively and consistently investigate the role of climatic shocks on internal migration for a large number of countries (64 countries). As highlighted in the conceptual framework explaining environmental drivers of migration from Black et al. (2011), migration is determined by the interactions among individual, meso and macro-level factors. Using individual educational attainment as a reference point to compare the importance of each factor underlying migration, we find that age and household size are key determinants of migration as well-documented in the literature (Borderon et al. 2019; Tsegai 2007). In addition, we find that relative to education and other individual characteristics including gender and marital status, climatic shocks also play an important role in influencing migration decisions.

Unlike traditional regression analysis which is typically driven by theoretical concepts to select which variables to focus on, in random forest models, drivers of migration are identified through a data driven approach (Best et al. 2020). Furthermore, as non-parametric models, random forest algorithms allow us to identify salient variables in large-scale and complex datasets without imposing any assumptions on the data distribution. Our finding that climatic shocks are as important as some individual characteristics in determining internal migration provides a solid evidence on the importance of environmental variables in influencing migration.

Focusing on the subnational regions where the climate variables have important explanatory power for migration, one consistent pattern that emerged was the positive relationship between a positive precipitation shock and migration across countries regardless of the level of GDP per capita. This is specifically the case for lower-income countries. Almost mirroring these results is the finding that migration decreases in the regions experiencing negative precipitation shocks, but this pattern is specific to regions located in countries in the low-income group. Our findings clearly lend support to the argument on the role of liquidity constraint on migration, that is worsening climatic conditions can reduce household income necessary for covering migration costs and consequently suppress potential migration (Cattaneo and Peri 2016; Cui and Feng 2020; Gray and Mueller 2012b; Gray and Wise 2016). When low-income countries experience higher precipitation than the average this may increase agricultural production and consequently generate extra income for travel and relocation costs. This pattern also appears to present in richer countries.

The major limitations for our study are related to migration data. First, relying on the migration-related question in the population censuses which asks the individuals about their previous place of residence, it is not possible to distinguish between short- and long-term migration and temporary and permanent migration. Previous studies have shown that different types of migration are adopted to cope with temperature and rainfall anomalies with temporary migration being used as an adaptive response for climate-vulnerable households (Bohra-Mishra et al. 2014; Joarder and Miller 2013; Mueller, Sheriff, et al. 2020; N. E. Williams and Gray 2020). Climate-related migration thus maybe underestimated in our case because we only capture migration that occurred within at least one year interval. Second, the number of administrative units varies considerably across countries. Movements in a country with a higher number of subnational regions are more likely to be classified as migration because it is naturally easier to cross a regional boundary. Although this affects our cross-national comparison of climate-related migration, the number of administrative regions is distributed randomly across income groups. Therefore, the bias in our comparison of climate-related migration across countries by their GDP per capita and the degree of aridity is limited. Third, our study is limited to 64 countries in the sample where countries in Asia (9) and Europe (6) are underrepresented. The findings therefore cannot be generalised as representing the climate-related migration patterns for the whole world.

Future extensions of this study could explore alternative machine learning methods, such as support vector machines that are well-suited for a truncation of high-dimensional feature space and combine them with inferential statistics like regression analysis. Following a similar study design, machine learning could be used to isolate relevant population samples on a national, regional, and individual level, which could then be analysed by the means of regression analysis. This could help further reduce the limitations that increasingly popular big social datasets impose on established multivariate regression analysis.

Despite the limitations, this is the first large-scale individual-level study that comprehensively and systematically assess the role of climatic factors on internal migration. The application of machine learning techniques allows us to identify important factors driving migration without imposing any theoretical nor statistical assumptions. This study provides consistent evidence supporting the argument of the previous literature on the role of economic resources which are influenced by climatic conditions in enabling mobility, which appears to be highly relevant also for internal migration.

Appendix

FIGURE A1: Positive Temperature Shock: Important drivers of migration identified by random forest models for 64 countries

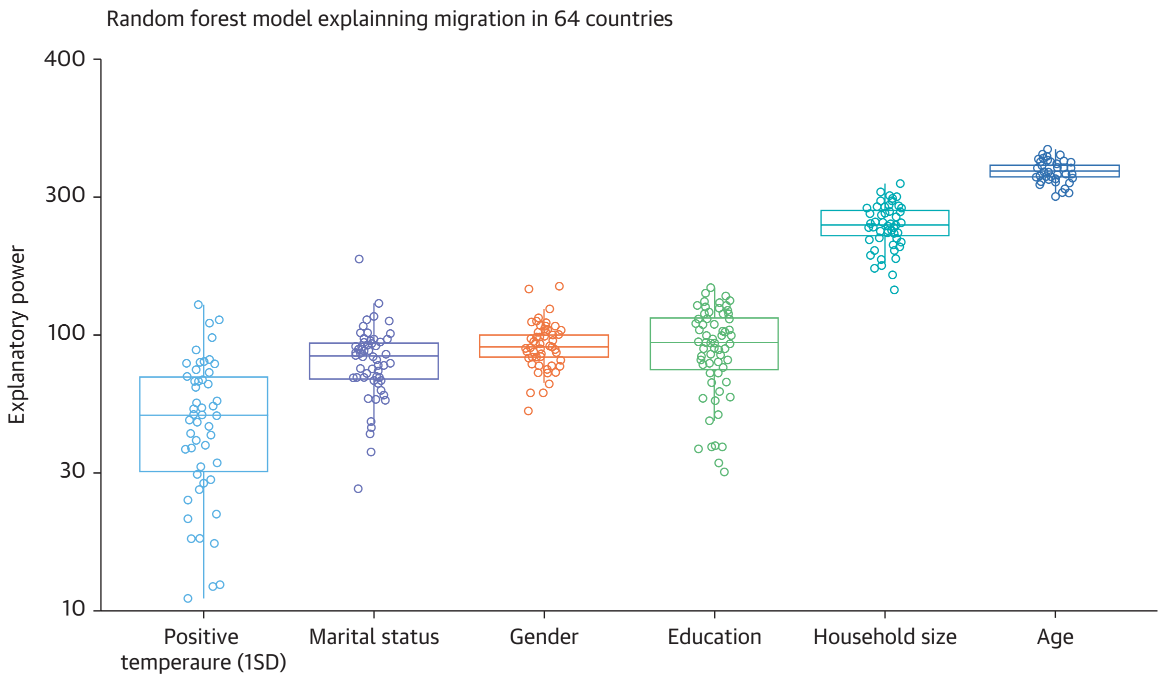
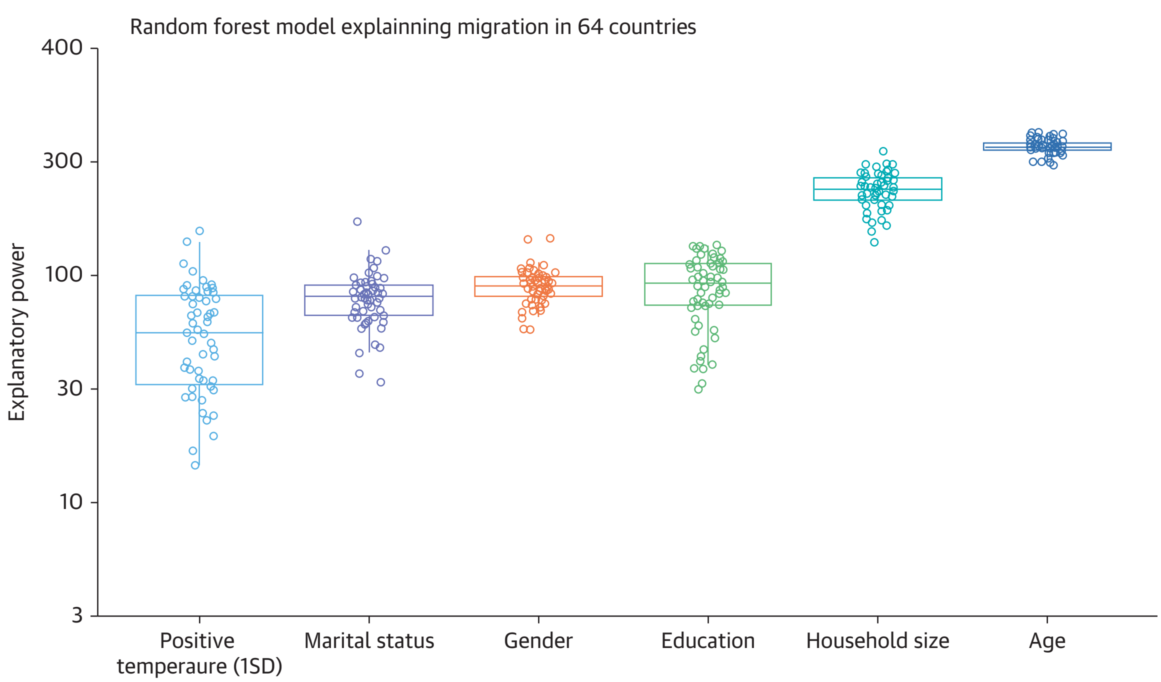


FIGURE A2: Positive Precipitation Shock: Important drivers of migration identified by random forest models for 64 countries



Raw data, including Table A1, are available in the online appendix:

https://doi.org/10.6084/m9.figshare.16780114.v1

TABLE A2: Regional comparison of changes in migration shares by various shocks and GDP per capita

Shock	GDP per Capita (log)	Group	N	Mean	SD	Share of regions
Negative Precipitation (1SD)	(0,6]	Less Migration	8	−0.03	0.02	36,36%
Negative Precipitation (1SD)	(0,6]	More Migration	14	0.04	0.05	63,64%
Negative Precipitation (1SD)	(6,8]	Less Migration	72	−0.03	0.02	64,86%
Negative Precipitation (1SD)	(6,8]	More Migration	39	0.04	0.03	35,14%
Negative Precipitation (1SD)	(8,10]	Less Migration	34	−0.07	0.04	47,22%
Negative Precipitation (1SD)	(8,10]	More Migration	38	0.04	0.05	52,78%
Negative Precipitation (1SD)	(10,Inf]	Less Migration	29	−0.07	0.04	50,88%
Negative Precipitation (1SD)	(10,Inf]	More Migration	28	0.06	0.02	49,12%
Positive Precipitation (1SD)	(0,6]	Less Migration	11	−0.04	0.03	14,10%
Positive Precipitation (1SD)	(0,6]	More Migration	67	0.04	0.03	85,90%
Positive Precipitation (1SD)	(6,8]	Less Migration	40	−0.03	0.02	36,70%
Positive Precipitation (1SD)	(6,8]	More Migration	69	0.04	0.03	63,30%
Positive Precipitation (1SD)	(8,10]	Less Migration	54	−0.05	0.05	48,21%
Positive Precipitation (1SD)	(8,10]	More Migration	58	0.06	0.04	51,79%
Positive Precipitation (1SD)	(10,Inf]	Less Migration	21	−0.05	0.04	35,59%
Positive Precipitation (1SD)	(10,Inf]	More Migration	38	0.07	0.03	64,41%
Positive Temperature (1SD)	(0,6]	Less Migration	5	−0.04	0.05	11,90%
Positive Temperature (1SD)	(0,6]	More Migration	37	0.04	0.04	88,10%
Positive Temperature (1SD)	(6,8]	Less Migration	57	−0.04	0.02	50,44%
Positive Temperature (1SD)	(6,8]	More Migration	56	0.05	0.04	49,56%
Positive Temperature (1SD)	(8,10]	Less Migration	41	−0.06	0.04	63,08%
Positive Temperature (1SD)	(8,10]	More Migration	24	0.05	0.05	36,92%
Positive Temperature (1SD)	(10,Inf]	Less Migration	25	−0.05	0.05	23,81%
Positive Temperature (1SD)	(10,Inf]	More Migration	80	0.06	0.02	76,19%

References

Adams, R. H. (1993). The economic and demographic determinants of international migration in Rural Egypt. *The Journal of Development Studies*, *30*(1), 146-167. https://doi.org/10.1080/00220389308422308

Agesa, J., & Agesa, R. U. (1999). Gender differences in the incidence of rural to urban migration: Evidence from Kenya. *The Journal of Development Studies*, *35*(6), 36-58. https://doi.org/10.1080/00220389908422601

Attewell, P., Monaghan, D. B., & Kwong, D. (2015). *Data Mining for the Social Sciences: An Introduction* (1st ed.). University of California Press. https://www.jstor.org/stable/10.1525/j.ctt13x1gcg. Accessed 12 September 2018

Aybek, C. M., Huinink, J., & Muttarak, R. (2015). Migration, Spatial Mobility, and Living Arrangements: An Introduction. In *Spatial Mobility, Migration, and Living Arrangements* (pp. 1-19). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10021-0_1

Barrios, S., Bertinelli, L., & Strobl, E. (2006). Climatic change and rural-urban migration: The case of sub-Saharan Africa. *Journal of Urban Economics*, *60*(3), 357-371. https://doi.org/10.1016/j.jue.2006.04.005

Beine, M., & Parsons, C. (2015). Climatic Factors as Determinants of International Migration. *The Scandinavian Journal of Economics*, *117*(2), 723-767. https://doi.org/10.1111/sjoe.12098

Bell, M. & Charles-Edwards, E. (2013). Cross-national comparisons of internal migration: An update on global patterns and trends. 10.13140/RG.2.1.3489.1769.

Bell, M., Charles-Edwards, E., Kupiszewska, D., Kupiszewski, M., Stillwell, J., & Zhu, Y. (2015). Internal Migration Data Around the World: Assessing Contemporary Practice. *Population, Space and Place*, *21*(1), 1-17. https://doi.org/10.1002/psp.1848

Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., & Kupiszewska, D. (2015). Internal Migration and Development: Comparing Migration Intensities Around the World. *Population and Development Review*, *41*(1), 33-58. https://doi.org/10.1111/j.1728-4457.2015.00025.x

Bernard, A., & Bell, M. (2015). Smoothing internal migration age profiles for comparative research. *Demographic Research*, *32*(1), 915-948. https://doi.org/10.4054/DemRes.2015.32.33

Bernard, A., Bell, M., & Charles-Edwards, E. (2014). Life-course transitions and the age profile of internal migration. *Population and Development Review*, *40*(2), 213-239. https://doi.org/10.1111/j.1728-4457.2014.00671.x

Best, K. B., Gilligan, J. M., Baroud, H., Carrico, A. R., Donato, K. M., Ackerly, B. A., & Mallick, B. (2020). Random forest analysis of two household surveys can identify important predictors of migration in Bangladesh. *Journal of Computational Social Science*. https://doi.org/10.1007/s42001-020-00066-9

Black, R., Adger, W. N., Arnell, N. W., Dercon, S., Geddes, A., & Thomas, D. (2011). The effect of environmental change on human migration. *Global Environmental Change*, *21, Supplement 1*, S3-S11. https://doi.org/10.1016/j.gloenvcha.2011.10.001

Bohra-Mishra, P., Oppenheimer, M., & Hsiang, S. M. (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences*, *111*(27), 9780-9785. https://doi.org/10.1073/pnas.1317166111

Borderon, M., Sakdapolrak, P., Muttarak, R., Kebede, E., Pagogna, R., & Sporer, E. (2019). Migration influenced by environmental change in Africa: A systematic review of empirical evidence. *Demographic Research (Special Collection)*, *11*, 491-544. https://doi.org/10.4054/DemRes.2019.41.18

Brockerhoff, M., & Eu, H. (1993). Demographic and socioeconomic determinants of female rural to urban migration in Sub-Saharan Africa. *The International Migration Review*, *27*(103), 557-577.

Cattaneo, C. (2007). *The Self-Selection in the Migration Process: What Can We Learn?* (No. 199). *LIUC Papers in Economics*.

Cattaneo, C., Beine, M., Fröhlich, C. J., Kniveton, D., Martinez-Zarzoso, I., Mastrorillo, M., et al. (2019). Human Migration in the Era of Climate Change. *Review of Environmental Economics and Policy*, *13*(2), 189-206. https://doi.org/10.1093/reep/rez008

Cattaneo, C., & Peri, G. (2016). The migration response to increasing temperatures. *Journal of Development Economics*, *122*(Supplement C), 127-146. https://doi.org/10.1016/j.jdeveco.2016.05.004

Chae, S., & Glick, J. E. (2019). Educational Selectivity of Migrants and Current School Enrollment of Children Left-Behind: Analyses in Three African Countries. *The International Migration Review*, *53*(3), 736-769. https://doi.org/10.1177/0197918318772261

Chindarkar, N. (2012). Gender and climate change-induced migration: proposing a framework for analysis. *Environmental Research Letters*, *7*(2), 025601. https://doi.org/10.1088/1748-9326/7/2/025601

Cui, X., & Feng, S. (2020). Climate Change and Migration. In K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics* (pp. 1-15). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57365-6_115-1

Curran, S. R., & Rivero-Fuentes, E. (2003). Engendering migrant networks: The case of Mexican migration. *Demography*, *40*(2), 289-307. https://doi.org/10.1353/dem.2003.0011

de Haan, A. (1997). Migration as family strategy: Rural-urban labor migration in India during the twentieth century. *The History of the Family*, *2*(4), 481-505. https://doi.org/10.1016/S1081-602X(97)90026-9

de Haas, H. (2011). *The determinants of international migration: Conceptualizing policy, origin and destination effects*. Oxford: International Migration Institute, IMI Working Paper.

Dillon, A., Mueller, V., & Salau, S. (2011). Migratory Responses to Agricultural Risk in Northern Nigeria. *American Journal of Agricultural Economics*, *93*(4), 1048-1061. https://doi.org/10.1093/ajae/aar033

Dustmann, C., & Glitz, A. (2011). Chapter 4 - Migration and Education. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 4, pp. 327-439). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00004-3

El-Hinnawi, E. (1985). *Environmental Refugees*. Nairobi: United Nations Environment Programme.

Findley, S. E. (1994). Does Drought Increase Migration? A Study of Migration from Rural Mali during the 1983-1985 Drought. *International Migration Review*, *28*(3), 539–553. https://doi.org/10.2307/2546820

Fussell, E., Hunter, L. M., & Gray, C. (2014). Measuring the environmental dimensions of human migration: The demographer's toolkit. *Global Environmental Change: Human and Policy Dimensions*, *28*, 182–191. https://doi.org/10.1016/j.gloenvcha.2014.07.001

Gemenne, F. (2011). Why the numbers don't add up: A review of estimates and predictions of people displaced by environmental changes. *Global Environmental Change*, *21, Supplement 1*, S41–S49. https://doi.org/10.1016/j.gloenvcha.2011.09.005

Ginsburg, C., Bocquier, P., Béguy, D., Afolabi, S., Augusto, O., Derra, K., et al. (2016). Human capital on the move: Education as a determinant of internal migration in selected INDEPTH surveillance populations in Africa. *Demographic Research*, *34*(30), 845–884. https://doi.org/10.4054/DemRes.2016.34.30

Gould, W. T. S. (1982). Education and internal migration: A review and report. *International Journal of Educational Development*, *1*(3), 103–111.

Gray, C., & Mueller, V. (2012a). Drought and population mobility in rural Ethiopia. *World Development*, *40*, 134–145. https://doi.org/10.1016/j.worlddev.2011.05.023

Gray, C., & Mueller, V. (2012b). Natural disasters and population mobility in Bangladesh. *Proceedings of the National Academy of Sciences*, *109*(16), 6000–6005. https://doi.org/10.1073/pnas.1115944109

Gray, C., & Wise, E. (2016). Country-specific effects of climate variability on human migration. *Climatic Change*, *135*(3–4), 1–14. https://doi.org/10.1007/s10584-015-1592-y

Hastie, T. (2009). Random forests. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (2nd ed., pp. 587–604). New York: Springer-Verlag. https://doi.org/10.1007/978-0-387-84858-7

Henderson, J. V., Storeygard, A., & Deichmann, U. (2017). Has climate change driven urbanization in Africa? *Journal of Development Economics*, *124*, 60–82. https://doi.org/10.1016/j.jdeveco.2016.09.001

Hoffmann, R., Dimitrova, A., Muttarak, R., Crespo Cuaresma, J., & Peisker, J. (2020). A meta-analysis of country-level studies on environmental change and migration. *Nature Climate Change*, 1–9. https://doi.org/10.1038/s41558-020-0898-6

IOM. (2014). *IOM Outlook on migration, environment and climate change*. Geneva: International Organization for Migration. https://publications.iom.int/system/files/pdf/mecc_outlook.pdf

Joarder, M. A. M., & Miller, P. W. (2013). Factors affecting whether environmental migration is temporary or permanent: Evidence from Bangladesh. *Global Environmental Change*, *23*(6), 1511–1524. https://doi.org/10.1016/j.gloenvcha.2013.07.026

Kanaiaupuni, S. M. (2000). Reframing the Migration Question: An Analysis of Men, Women, and Gender in Mexico. *Social Forces*, *78*(4), 1311–1347. https://doi.org/10.2307/3006176

Kuhn, R. (2005). *The Determinants of Family and Individual Migration: A Case-Study of Rural Bangladesh* (No. POP2005- 05). Boulder, CO: IBS Population Program, University of Colorado. https://ibs.colorado.edu/pubs/pop/pop2005-0005.pdf

Lin, M., Lucas, H. C., & Shmueli, G. (2013). Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. *Information Systems Research*, *24*(4), 906–917. https://doi.org/10.1287/isre.2013.0480

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*(3), 355–374. https://doi.org/10.1016/j.jeem.2012.02.001

Massey, D. S., Alarcón, R., Durand, J., & González, H. (1987). *Return to Aztlan: The Social Process of International Migration from Western Mexico*. Berkeley; Los Angeles; London: University of California Press. http://www.jstor.org/stable/10.1525/j.ctt1ppp3j. Accessed 17 April 2017

Massey, D. S., & Espinosa, K. E. (1997). What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis. *American Journal of Sociology*, *102*(4), 939–999.

Mastrorillo, M., Licker, R., Bohra-Mishra, P., Fagiolo, G., D. Estes, L., & Oppenheimer, M. (2016). The influence of climate variability on internal migration flows in South Africa. *Global Environmental Change*, *39*(Supplement C), 155-169. https://doi.org/10.1016/j.gloenvcha.2016.04.014

Minnesota Population Center. (2019). Integrated Public Use Microdata Series, International: Version 4.2 [dataset]. University of Minnesota. Minneapolis, MN. https://doi.org/10.18128/D020.V7.2

Mueller, V., Gray, C., & Hopping, D. (2020). Climate-Induced migration and unemployment in middle-income Africa. *Global Environmental Change*, *65*, 102183. https://doi.org/10.1016/j.gloenvcha.2020.102183

Mueller, V., Gray, C., & Kosec, K. (2014). Heat stress increases long-term human migration in rural Pakistan. *Nature Climate Change*, *4*(3), 182-185. https://doi.org/10.1038/nclimate2103

Mueller, V., Sheriff, G., Dou, X., & Gray, C. (2020). Temporary migration and climate variation in eastern Africa. *World Development*, *126*, 104704. https://doi.org/10.1016/j.worlddev.2019.104704

Myers, N. (1993). Environmental refugees in a globally warmed world. *BioScience*, *43*(11), 752-761. https://doi.org/10.2307/1312319

Myers, N. (2002). Environmental refugees: A growing phenomenon of the 21st century. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1420), 609-613. https://doi.org/10.1098/rstb.2001.0953

Phuong, N. T., Tam, T. N. T. M., Nguyet, N. T., & Oostendorp, R. (2008). *Determinants and Impacts of Migration in Vietnam* (No. 01). *Working Papers*. Development and Policies Research Center (DEPOCEN), Vietnam. https://ideas.repec.org/p/dpc/wpaper/0108.html. Accessed 5 April 2021

Piguet, E. (2010). Linking climate change, environmental degradation, and migration: a methodological overview. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(4), 517-524. https://doi.org/10.1002/wcc.54

Piguet, E., Kaenzig, R., & Guélat, J. (2018). The uneven geography of research on "environmental migration." *Population and Environment*, *39*(4), 357-383. https://doi.org/10.1007/s11111-018-0296-4

Quinn, M. A., & Rubb, S. (2005). The importance of education-occupation matching in migration decisions. *Demography*, *42*(1), 153-167.

Reed, H. E., Andrzejewski, C. S., & White, M. J. (2010). Men's and women's migration in coastal Ghana: An event history analysis. *Demographic Research*, *22*(25), 771-812. https://doi.org/10.4054/DemRes.2010.22.25

Rigaud, K. K., Sherbinin, A. de, Jones, B. R., Bergmann, J. S., Clement, V. W. C., Ober, K. J., et al. (2018). *Groundswell: Preparing for Internal Climate Migration* (No. 124719) (pp. 1-256). The World Bank. http://documents.worldbank.org/curated/en/846391522306665751/Main-report. Accessed 3 April 2019

Rogers, A., & Castro, L. J. (1981). *Model Migration Schedules*. *IIASA Research Report* (Vol. 81). Laxenburg, Austria.

Rosenzweig, M. R., & Stark, O. (1989). Consumption Smoothing, Migration, and Marriage: Evidence from Rural India. *Journal of Political Economy*, *97*(4), 905-926.

Simpson, N. B. (2017). Demographic and economic determinants of migration. *IZA World of Labor*, (373). https://doi.org/10.15185/izawol.373

Sobek, M. (2016). Data prospects: IPUMS-International. In M. J. White (Ed.), *International Handbook of Migration and Population Distribution* (pp. 157-174). New York: Springer Netherlands. //www.springer.com/gb/book/9789401772815. Accessed 12 September 2018

Thiede, B. C., Gray, C., & Mueller, V. (2016). Climate variability and inter-provincial migration in South America, 1970-2011. *Global Environmental Change*, *41*, 228-240. https://doi.org/10.1016/j.gloenvcha.2016.10.005

Tsegai, D. (2007). Migration as a Household Decision: What are the Roles of Income Differences? Insights from the Volta Basin of Ghana. *The European Journal of Development Research*, *19*(2), 305-326. https://doi.org/10.1080/09578810701289212

Williams, N. (2009). Education, gender and migration in the context of social change. *Social science research*, *38*(4), 883-896.

Williams, N. E., & Gray, C. (2020). Spatial and Temporal Dimensions of Weather Shocks and Migration in Nepal. *Population and environment*, *41*(3), 286-305. https://doi.org/10.1007/s11111-019-00334-5

Willmott, C. J., & Matsuura, K. (2015). Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1900 - 2014). http://climate.geog.udel.edu/~climate/html_pages/Global2014/README.GlobalTsP2014.html. Accessed 10 January 2021

GWSP
GLOBAL WATER
SECURITY & SANITATION
PARTNERSHIP

WORLD BANK GROUP

SKU W21045