A Reappraisal of the Migration-Development Nexus

Testing the Robustness of the Migration Transition Hypothesis

Nicolas Berthiaume
Naomi Leefmans
Nienke Oomes
Hugo Rojas-Romagosa
Tobias Vervliet

WORLD BANK GROUP

Macroeconomics, Trade and Investment Global Practice
January 2021

Abstract

This paper tests the migration transition hypothesis that emigration flows first increase and later decrease with a country's economic development. Using a migration version of the gravity model, this hypothesis is tested on a global panel data set comprising 180 origin and destination countries and a 50-year timeframe (1970–2020). This is the most extensive panel data set used so far to test the migration transition hypothesis. The results confirm the existence of an inverted U-shaped relationship between development and emigration within a cross-country panel setting. Nevertheless, the migration hump cannot be interpreted as a causal relationship: for a given low-income country, an increase in economic development is not found to lead to higher emigration. For a subsample of 44 countries that have transitioned from low-income to middle-income status, emigration has rather declined with economic development. The migration transition hypothesis is therefore unfounded. Instead, the migration hump appears to be driven by an underlying cross-sectional pattern that cannot be fully controlled: middle-income countries tend to exhibit higher emigration rates than low- or high-income countries. The findings of this paper have important policy implications: development programs can simultaneously promote economic development and reduce emigration.

A Reappraisal of the Migration-Development Nexus: Testing the Robustness of the Migration Transition Hypothesis

Nicolas Berthiaume[a], Naomi Leefmans[b], Nienke Oomes[a],

Hugo Rojas-Romagosa[c] and Tobias Vervliet[a]

[a] SEO Amsterdam Economics

[b] University of Amsterdam

[c] World Bank

1   Introduction

Globalization has facilitated physical mobility and as a result enabled international migration to increase from 92 million in 1960 to 244 million in 2017.[1] The traditional view that the root cause of these rising migration flows has been a lack of economic development in origin countries has resurfaced in the past few years within the policy debates of both sending and receiving countries.[2] Would-be migrants, the argument goes, decide to move primarily in search of higher wages and income abroad. In this framework, exogenous non-economic factors such as natural disasters and conflicts at origin are secondary.

The direct relation between income differentials and emigration originates from the neoclassical theory of migration.[3] This theory posits that a higher domestic reservation wage reduces the relative expected returns on emigration, as opposed to staying at home. This implies that, the larger the income and wage differentials between countries, the higher the migration pull factors are. Consequently, emigration is predicted to decrease as income gaps between origin and destination countries close.

An important policy implication of this theory is that high-income countries can decrease immigration through policies that help low-income countries raise their average incomes and development levels. When income differentials decline as a result, so will the migration flows from low-income to higher-income countries. This will also relieve strained borders and stem the brain drain that negatively affects developing countries (Caselli, 2019). Accordingly, since the 1990s, policy makers, academics and development NGOs have advocated a triad of policies aimed at fostering development in emigration countries through aid, trade liberalization, and temporary and return migration (De Haas, 2007).

However, although these models are intuitively appealing, they do not adequately explain observed patterns of migration. Empirical evidence shows that migration determinants do not depend only on economic factors such as income and wages, but also on migrant networks abroad, foreign immigration policies, and demographic transitions (Clemens, 2014). The migration transition hypothesis developed in Zelinsky's (1971) seminal paper, on the other hand, accounts for both these economic and non-economic migration determinants. This creates a richer interrelation between migration and economic income levels. In particular, this hypothesis predicts a nonlinear inverted-U relationship between development and migration. Emigration first rises as development increases in a given origin country, until a so-called migration transition turning point is reached, after which emigration starts declining. As explained in Clemens (2014), this phenomenon can be explained by factors such as, among others, rising inequality, gradually relieving credit constraints, and structural labor market changes leading to worker dislocation, which might all accompany the economic development process.

There is an extensive literature on the determinants of migration that has tested Zelinsky's hypothesis. Using cross-section data many studies find the inverted U-shaped relationship between levels of GDP per capita and the share of emigrants, even after controlling for other determinants of migration (Djajic *et al.*, 2016; Dao *et al.*, 2018; Idu, 2019). However, testing for a migration hump using cross-section

---

[1] These values were computed using the World Bank's Global Bilateral Migration (Özden *et al.*, 2011) and the United Nations' Trends in International Migrant Stocks (UN Department of Economic and Social Affairs, 2017) databases. This corresponds to an increase in migration flows from 3% of the World population in 1960 to 3.2% in 2017.
[2] This notion was first put forth in Todaro (1969) and Lucas (1975), and is exemplified in the European Commission's (EC) European Agenda for Migration (EC, 2015), for instance.
[3] The neoclassical model of migration was first elaborated in Ravenstein (1985). See De Haas (2011) for a survey on the different theories on the determinants of migration.

data leaves important considerations unaccounted for, such as reverse causality and the migration transition's longitudinal dimension, as the transition takes place over an extended time period in a given origin country. Other studies have tested for a hump shape using panel data (Mayda, 2010; Bertoli and Huertas-Moraga, 2013). However, these papers use a limited number of country-time points, which restricts the empirical strength of their results. Other papers test the inverted-U relationship using solely migration flows to OECD destinations (Lull, 2016; Benček and Schneiderheinze, 2019). These studies, however, exclude the possibility that migrants from low-income countries can also migrate to other low- or medium-income countries. Since the average share of migration from all origins to non-OECD destinations is 50% over the 1960-2017 period,[4] we include such migration flows in order to incorporate all migration corridors in the analysis.

The aim of this paper is to test for the inverted U-shape between emigration and development using a large panel database. We employ a comprehensive global panel data set with 180 origin and destination countries on a 50-year timeframe (1970-2020).[5] This allows us to empirically test for bilateral migration dynamics not only across countries but also across time with a relatively large number of observations. Because of its large longitudinal dimension, it is well suited for testing the migration transition hypothesis' central prediction, which is a long-run phenomenon per origin country (De Haas, 2010). Our empirical specification is based on the random utility-maximization (RUM) model, which provides the micro-foundations for a migration version of the gravity model.[6] We employ a gravity-migration specification with a large number of fixed effects, which control for several observed and unobserved origin-, destination-, time- and country-pair-specific characteristics deemed to influence migration. We introduce both a linear and a squared GDP per capita at origin term (our proxy for development levels) to test for the non-linear inverted-U shape. This term is instrumented using its period-to-period lag in order to tackle reverse causality. The data set presented in this paper further contributes to Llull (2016), who employs a similar panel data set including bilateral migration flows for the 1960-2000 time period but does not test for the inverted U-shaped relationship between development and emigration.

To our knowledge, this is the first paper on the migration transition hypothesis that tests the RUM model on a global panel data set that extends over a period of 50 years and includes bidirectional flows for 180 origin and destination countries. This comprehensive database accounts for all potential migration flows, and not merely flows to OECD destinations. As stated above, about half of all international migration, on average, was to non-OECD destinations. Merely including OECD destinations would therefore leave out a large portion of all migration flows. Furthermore, we reduce the bias due to the presence of zeros in the dependent variable using a Poisson Pseudo-Maximum-Likelihood estimator with High-Dimensional Fixed Effects (PPML-HDFE), and not by simply omitting them or resorting to data aggregations. Lastly, we conduct additional alternative tests of an inverted U-shaped relationship, while previous studies have generally merely run quadratic model estimations and hence ran into the risk of incorrectly finding an extremum.

---

[4] Computed using the World Bank's Global Bilateral Migration (Özden *et al.*, 2011) and the United Nations' Trends in International Migrant Stocks (UN Department of Economic and Social Affairs, 2017) databases.

[5] Data on international migrant stocks in 2019 is used as a proxy for 2020.

[6] Gravity models are more commonly employed in the trade literature, but several migration studies also use them. See Beine *et al.* (2016) for an extensive review of the migration literature employing RUM micro-founded gravity models.

Our results confirm the existence of an inverted U-shaped relationship between development and emigration within a cross-country (panel) setting. This result is robust to the inclusion of additional control variables and the estimation of the empirical model on alternative time subsamples. It is also robust to the inclusion of an interaction term between geographical distance and income at origin, and several additional tests for the existence of the inverted-U relationship.

However, we cannot conclude that our findings yield evidence of a causal link between development at origin and emigration flows. The reason is that multilateral resistance to migration (i.e., that the attractiveness of a given country depends on the latent attractiveness of other potential destinations) is not fully accounted for. The only viable way to adequately correct for this is to also include origin-time fixed effects next to other (origin, destination-time, time, and country-pair) fixed effects that we do include. However, like all other papers in the existing literature on this topic, our econometric model does not allow for the inclusion of origin-time fixed effects as these would be perfectly collinear with our origin-time-varying variable of interest: GDP per capita at origin. With this endogeneity issue remaining unsolved we cannot claim that our results establish a causal relationship.

We perform several robustness analyses to test whether an initial increase in economic development indeed leads to higher emigration. To this end, we test, for a subsample of countries that have actually transitioned from the low-income to the middle-income category, whether their emigration has increased with development, by applying both a linear and a quadratic version of our regression model. From this and several other robustness tests, we do not find that the inverted-U relation between development and emigration based on panel data also implies such a relation for an individual low-income country over time. Accordingly, drawing the conclusion that the inverted-U relationship between economic development and emigration is causal seems unfounded.

Several authors (e.g. De Haas, 2019, Clemens and Postel, 2018) have concluded from the migration transition hypothesis that as low-income countries develop, their emigration will tend to increase before declining after the turning point and that development aid is therefore not a proper instrument to reduce emigration from low-income countries. Our findings do not imply this conclusion. On the contrary, for a subsample of countries that transitioned from low to middle-income (excluding China and India), we find that, as low-income countries develop economically, their emigration actually declined. This obviously has important policy implications for development cooperation: it suggests that development programs can actually reduce emigration from low-income countries if they are successful at promoting local economic development.

The remainder of this paper is structured as follows. In section 2, we review the theories that might give grounds to the existence of a migration-development inverted U-shaped 'life cycle' in any given country, as well as the current empirical evidence for them. Section 3 describes the data we use and provides a descriptive analysis. Section 4 outlines our empirical methodology and section 5 presents our results, also including several robustness analyses. Section 6 concludes.

2   The migration-development 'life cycle'

This section presents a literature review on the migration transition theories as well as the existing empirical evidence of the inverted U-shaped relationship between development and migration.

2.1 Theory

The migration transition hypothesis (Zelinsky, 1971; Gould, 1979) sustains that economic, demographic, and socio-political forces, which co-occur with development, might also influence migration decisions. Under certain assumptions, such factors can jointly explain an inverted U-shaped relation between migration and development levels.

Following De Haas (2010), these factors affecting migration decisions can be grouped into migration capabilities and migration aspirations.

On the one hand, migration capabilities (MC) can be expected to monotonically increase with development indicators such as income and education, as well as with the creation of migrant networks abroad. First, income growth implies that potential migrants are better able to finance migration (Vanderkamp, 1971; Faini and Venturini, 2010). This effect can be compounded by the impact of remittances from migrant communities abroad. Second, improvements in education and human capital raise the number of feasible migration destinations by increasing the number of visa classes (which are usually skilled-employment work visas) that migrants can obtain (Flahaux and De Haas, 2016; Ortega and Peri, 2013). Third, would-be migrants' relationships with previous migrants already abroad may improve their ability to integrate in a given destination country, thereby further increasing migration capabilities (Massey, 1988). Yet, when the migrant population abroad grows, the positive network externalities generated by it may eventually disappear, due to the formation of a localized culture, gradually eroding the link between the established foreign network and potential domestic migrants (Epstein, 2008). Overall, with development, the rise in disposable income, human capital levels and migrant communities abroad leads to an increase in capabilities to emigrate. These MC can be expected to start growing more and more rapidly at first, because of the compounding impact of migrant networks and remittances, and later decelerating due to the formation of a localized culture with decreasing links with potential migrants in origin countries. This initial acceleration and later deceleration of migration capabilities with development is shown as the S-shaped MC curve in Figure 1.
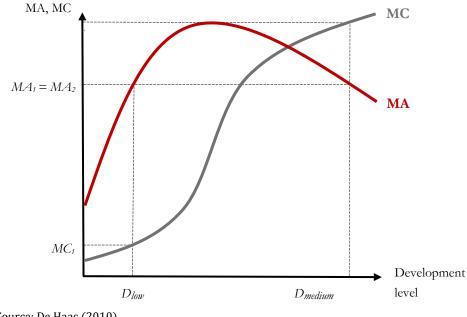
On the other hand, migration aspirations (MA) are more likely to have an inverted U-shape. Migration aspirations are a function of several factors, all of which are likely to first rise and later decrease with a country's economic development (Clemens, 2014). These factors include:

(i)   Population growth initially increases with development due to declining mortality rates, and at some point starts decreasing with further development due to declining fertility rates. The initial increased population growth generates labor market pressures at home and thus increases demand for emigration, while at some point reduced population growth reduces emigration aspirations (Zelinsky, 1971).

(ii)   Opportunity costs of migration for capital owners initially decrease with development and stop falling once the relative prices of production factors have adjusted to the economy's opening to international trade (Samuelson, 1948[7]; Martin and Taylor, 1996).

---

[7] According to the Stolper-Samuelson theorem, in a relatively poor country with an abundance of labor, trade liberalisation will increase the exports and relative price of the labor-intensive good and decrease the price of the capital-intensive goods. This is translated into a more than proportional increase in labor wages and a simultaneous reduction in capital returns. For capital owners, opportunity costs of migration (which include these foregone capital gains at home) are thereby reduced, increasing migration incentives.

(iii) Rising domestic inequality with development which can, for some subset of the population, increase the gap between expected and actual income, leading to an initial rise in migration aspirations (Stark, 2006).[8] Once the subset of the population with the highest gap between expected and actual income has migrated, this gap is on average reduced in the total population, causing a fall in aggregate migration aspirations.[9]

Figure 1 illustrates the hump-shaped line for migration aspirations and the S-shaped curve for migration capabilities. At development levels D_{low} and D_{medium}, we assume that one's aspiration to migrate is the same, at $MA_1 = MA_2$. Yet migration capabilities at D_{low} are much lower than at the higher development stage D_{medium}. For an equal aspiration to migrate, this difference in capabilities is expected to be the reason why poorer individuals tend to migrate less. Conversely, possessing both a strong willingness to migrate and sufficient capabilities to act upon it, medium earners are most likely to emigrate. On the other hand, since high-income individuals possess the required ability but lack the willingness to migrate, their propensity to do so will be lower.

Figure 1   The migration transition hypothesis at the individual level

Source: De Haas (2010).

At the country level and over time, we therefore expect emigration to first rise as domestic development rises, until a certain 'turning point' at which migration aspirations and capabilities are both relatively high. From this point onwards, capabilities grow just marginally with development, while migration aspirations fall, gradually pulling aggregate emigration rates downwards. Migration transition theories

[8] As domestic inequality rises, so does the income gap between the lower and higher ends of the income distribution. This lowers relative income for the poorest, and thus raises income expectations. Since migrating abroad may be a way to achieve this new level of expected income due to inter-country income differences, this can foster migration aspirations at the lower end of the income distribution.

[9] In reality, this phenomenon generally does not generate a clear inverted U-shaped relationship between development and migration aspirations. Inequality in a country can rise and fall more than once as development increases. Nevertheless, inequality has a clear impact on the gains from migration attained by workers at different points in the income distribution and in time: as inequality rises, migration aspirations are thought to increase in tandem, and vice-versa (Borjas, 1987).

therefore collectively predict that emigration has an inverted U-shaped 'life cycle' that is a function of the stage of development in the source country (Hatton and Williamson, 2011).

2.2 Empirical evidence

The inverted U-shaped relationship between migration and development has recently been observed in cross-sectional nonparametric regressions (Clemens, 2014; Dao, Docquier, Parsons and Peri, 2018). The turning point is graphically found to lie at a gross domestic product (GDP) per capita level varying from $ 4,000 to around $10,000 (in 2019 US dollars, adjusted for purchasing power parity (PPP)). Countries with medium levels of development are associated with the highest emigration rates, while both underdeveloped and highly developed countries exhibit comparatively low rates of emigration. Clemens (2014) and Dao *et al.* (2018) report that both the (initially) positive and the (later) negative relationships between emigration and GDP per capita levels were statistically significant. Clemens (2014) found that this cross-sectional, hump-shaped association holds for every decade since 1960 and becomes more pronounced with time. The turning point in GDP per capita remains at the same level, whereas the corresponding emigration rate increases over time. De Haas (2010) showed that the same cross-sectional inverted U-shaped relationship holds when using the human development index (HDI) instead of GDP per capita values.

It is not sufficient to merely observe that migration traces an inverted U-shaped pattern with development for a given year across countries. There are a number of studies that test for the existence of the migration hump using parametric regressions in such a cross-sectional setup, such as Djajic *et al.* (2016), Dao *et al.* (2018) and Idu (2019). However, this leaves at least three important considerations unaccounted for:

First, the migration transition hypothesis' central prediction is that this relationship ought to hold on average over time in any given country, and not merely in a given year across countries. That is, it is expected to hold in the longitudinal rather than in the cross-sectional dimension (Hatton and Williamson, 2011).

Second, development can be expected to affect migration flows, but rising migration also affects development levels, for instance through the remittances it generates. This can lead to reverse causality problems, which cannot be adequately tackled in a cross-sectional set-up.

Third, it can be expected that migration decisions strongly depend on observed or unobserved idiosyncratic characteristics of origin and destination countries. Examples of these factors are migration policies or individual preferences for migration, or drivers affecting pairs of countries, such as geographical distance or linguistic proximity. It is important to consider and correct for all costs and benefits related to every possible migration channel available to a would-be migrant.

Other studies have investigated the relationship between development at origin and emigration using panel data. Although not specifically testing the migration transition hypothesis, these authors have included a squared term in their specifications to test for nonlinearities in the migration-development nexus. These studies nonetheless use an insufficient number of country-time points to adequately test for this relationship. Mayda (2010) focusses on flows from 79 origins countries to 14 OECD destinations, and therefore does not incorporate other types of flows (e.g. South-South) in the analysis. The data only contains migration observations for 15 years (1980-1995). Similarly, Bertoli and Huertas-Moraga (2013) test their migration model on a 12-year timeframe (1997-2009) for 61 origins to a single destination.

One paper that employs a similar methodology to ours is the paper by Llull (2016). His paper exploits a relatively new database of bilateral migrant stocks and finds heterogeneous effects of income gains on migration prospects depending on distance. Like our paper, he uses a gravity-migration specification which is tested using panel data. Moreover, Llull (2016) employs a similar bilateral data set although the data we present in this paper is more temporally extended.

Despite the similarities, this paper differs from Llull (20106) in three important ways. First, Llull (2016) does not test for the existence of a hump-shaped relationship between emigration and development. Second, he uses migrant stocks instead of migration flows as the dependent variable, which is not in line with the specification's micro-foundation (Beine *et al.*, 2016). Third, Lull (2016) does not use the PPML-HDFE technique and instead employs the Ordinary Least Squares (OLS) technique. OLS is known not to perform well when the proportion of zeros in the dependent variable is high, which is the case here. It also yields relatively high biases in the presence of heteroscedasticity (Silva and Tenreyro, 2006, 2011).

A second, and as far as we know the only other, paper that is similar to ours is Benček and Schneiderheinze (2019), who more recently tested systematically for the existence of the migration hump. They find a negative relationship between income and emigration that is independent from the origin country's initial income level. Similar to this paper, they investigate the existence of the hump shape not only in cross-section but also over time.

Our methodology and data differ from Benček and Schneiderheinze (2019) in three ways. First, we explore all bilateral migration flows, whereas Benček and Schneiderheinze (2019) only focus on unilateral emigration flows to OECD countries. Second, we employ an estimation method owing to which we are able to limit the estimation bias due to the large number of zeros in our migration flow variable without having to exclude these observations. We do not make such sample selections as it might generate bias due to the exclusion of many potential destination countries. Third, we include a complete set of origin- and destination-time fixed effects, which reduces, although not fully eliminates, the potential endogeneity issues.

3   Data and descriptive analysis

3.1   Data

For the empirical analysis, we compiled an extensive panel data set comprising bilateral migration flows of 180 origin and destination countries for each decade from 1970 to 2020 (using 2019 as a proxy for 2020).

The dependent variable is the bilateral migration flow in each of the five decades from 1970 to 2020. Each of the explanatory variables we include in our model are varying in the origin-country and time dimensions only. We also employ fixed effects that vary in the destination, time and country-pair dimensions. All are averaged over decades, from $t-10$ to $t-1$. The dependent variable under study in our analysis is the decadal bilateral migration flow for the 1970-2020 time period.[10] Following Beine and Parsons (2015), migration flows are computed as the decade-to-decade difference in stocks, where, if M_{ijt} represents the stock of migrants from country i living in destination j at time t, the migration flow in period t is defined as $m_{ijt} = M_{ijt} - M_{ij,t-1}$.

---

[10] We thus use the following reference years: 1970, 1980, 1990, 2000, 2010, and 2020.

To measure this variable, we merge two migrant stock databases produced by the World Bank and the UN. For the years 1960-2000, we use the World Bank's Global Bilateral Migration database compiled by Özden *et al.* (2011). This is based on raw data from the Global Migration Database of the United Nations Department of Economic and Social Affairs of the Population Division (UN DESA, 2008). It contains migrant stock data by country of origin compiled from a collection of 3,500 censuses spanning 230 migrant destinations, for every decade from 1960 to 2000.[11] For the years 2010 and 2020 we combine this World Bank database with the Trends in International Migrant Stocks data from UN DESA (2019), which contains data for the following reference years: 1990, 1995, 2000, 2005, 2010, 2015 and 2019. This methodology is akin to Özden *et al.* (2011). The year 2019 is used as a proxy for the year 2020.

In these databases, migrants are defined as foreign-born individuals who have moved to a different country.[12] As explained in Özden *et al.* (2011) and UN DESA (2019), this has advantages over defining them by their citizenship. The latter definition does not provide a consistent measure of international migrant stocks because of differing citizenship laws across nations, and because people in some countries can acquire citizenship after having been a migrant for a number of years. This definition better captures the concept of migration as a "movement of a person or a group of persons, either across an international border, or within a State" (International Organization for Migration, 2011).

Both databases are based on the same underlying migration data and share many of the same processing methods. In both cases, the UN's Population Division census data is used to compile the database. The same country list is employed for both databases, although the UN DESA data contains six more countries than the World Bank's. In our merged data set, we only count those countries included in both databases. The original data suffers from a substantial amount of missing observations because many countries do not release national census data every 10 years. These may be prohibitively expensive in terms of labor intensity, can be abandoned because of exogenous factors, such as civil unrest or conflict, or are never released for political reasons. The authors chose to minimize the number of gaps in the data through interpolations. For the 'in-between' years (1970, 1980, 1990 for the 1960-2000 World Bank data and 2000 and 2010 for the UN DESA data), they do so by assuming a linear trend before and after missing data points. Where data are lacking for the beginning or end decades, they use growth rates in migration, taken from the UN Total Migrant Stock database (2006), to estimate bilateral migrant stocks. It is important to note, however, that since both databases use interpolations and predictions to fill in for missing values, our compiled bilateral database will also include a number of predicted values.[13] As a result, our estimation results are partially based on using predicted values as independent variables, which leads to increased uncertainty on the results.

---

[11] During the timeframe covered by these censuses, many regions reshaped their political boundaries, such as the USSR and Germany. For this reason, authors define their "master" country list as the most current set of countries.

[12] This definition is used where possible. Whenever birthplace information is missing, the authors identify international migrants using the citizenship criterion in order to minimize the amount of missing data points.

[13] For the World Bank database, for example, Özden *et al.* (2011) report that around 30% of countries have no missing data, 60% have one to three missing census rounds and the remaining 10% have four to five missing rounds. However, the countries with no missing data represent 68% of total world migration, while countries with just one or two missing rounds represent an additional 22%. Hence, 90% of world migration in this database is either based on raw data or by interpolating one or two data points of a total of five.
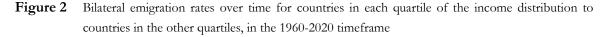
The UN DESA (2019) database differs from the World Bank database (Özden *et al.*, 2011) in two ways. Firstly, the UN DESA (2019) also adds data on refugees if available. Secondly, UN DESA (2019) used nationally representative surveys to complement the international migrant stock estimates based on population censuses and registers used in both databases.

We follow Rojas-Romagosa and Bollen (2018) by appending the data sets using the most recent UN international migrant stock data for the year 2019. Employing decadal data enables us to closely map our data to the population census rounds, which are done every decade. As in Beine and Parsons (2015), we set negative flow values to zero.

To our knowledge, this is the most extensive panel data set used so far in the literature to test for the existence of the migration hump. First, the large time dimension (50 years) has not been used to test the migration transition hypothesis before and it is well-suited to capture migration's long-run dynamics. Second, the large set of origins and destinations (180 countries, see Appendix Table A. 1) enables us to test the model on every possible migration direction, and not just South-North flows. Appendix Table A.2 contains the definitions and sources for all variables used in this paper.

3.2   Descriptive statistics

Table A. 4 in the Appendix shows the summary statistics for our dependent variable of interest (migration flows), migration rates (migrant stocks over population), our explanatory variable of interest

Figure 2   Bilateral emigration rates over time for countries in each quartile of the income distribution to countries in the other quartiles, in the 1960-2020 timeframe

(GDP per capita) and all other explanatory variables used in this study. Notably, with their highly positive skewness, the migration flow and rate distributions are heavily skewed towards the left. This reflects the large number of migration directions with small or zero flows of migrants.[14] The share of migration to OECD countries is equal to around 50% on average.[15]

The evolution of bilateral migration over time from each income quartile of the distribution of GDP per capita country-time points in our data set to all other quartiles is depicted in Figure 2. A similar graph using the World bank's classification of countries into low- lower middle- upper middle- and high-income groups presenting bilateral emigration rates over time for these income groups to all other income groups can be found in Figure A.1. in the appendix. As shown by both figures, migration rates are generally highest for lower- and upper-middle-income countries than for low- and high-income countries for each time period shown, in line with the cross-sectional migration hump.

4 Empirical analysis: Methodology

4.1 The canonical RUM model

The Random Utility-Maximization (RUM) model has recently been used in the migration literature, see Beine *et al.* (2016). This approach allows us to rigorously micro-found a migration version of the gravity model that is more commonly employed in the trade literature since Tinbergen's (1962) seminal contribution. The RUM expression of the location-decision problem faced by a would-be migrant (which translates into a simple utility-maximization problem) includes country-pair-specific utility components which call for the inclusion of bilateral (gravity) variables into the empirical model. Let us consider the location-decision problem faced by an individual h that considers migrating from a given country i to country j at time t. RUM models describe the utility derived from this move as:

$$U_{hijt} \equiv w_{ijt} - c_{ijt} + \theta_{hijt}, \tag{1}$$

where w_{ijt} denotes a deterministic component of utility and c_{ijt} represents the cost of migrating from i to j at time t. These can both be modelled as a function of observable variables, which should capture anything increasing or reducing the attractiveness of a particular destination and should include location- or country-pair-specific elements (Bertoli and Huertas-Moraga, 2013).

Conversely, since θ_{hijt} is an individual-specific stochastic term, it cannot be observed. As has been repeatedly done in the migration literature, we assume that θ_{hijt} follows an independent and identically

---

[14] To be precise, 153,700 migrant flow observations are equal to zero, or about 40.95% of the total.

15 This was computed using our country-time data set as the flow of international migrants from all possible origins having moved to an OECD destination in a given decade, averaged over the entire time sample.

distributed extreme value type 1 distribution à la McFadden and Zarembka (1974). Applied to equation (1), the expected share of individuals residing in *i* who move to *j* at time *t*, $E(p_{ijt})$, can then be written as:

$$E(p_{ijt}) = \frac{e^{w_{ijt}-c_{ijt}}}{\sum_{l \in D} e^{w_{ilt}-c_{ilt}}},$$ (2)

where D is the set of all countries the individual can choose from, l represents any country in this choice set, and $p_{ijt} \in [0,1]$ is the actual share of share of individuals residing in *i* who move to *j* at time *t* . By definition, the expected scale of the migration flow from country *i* to country *j* at time *t* is $E(m_{ijt}) = E(p_{ijt})s_{it}$, where s_{it} represents the size of the population residing in country *i* at time *t*. We can thus re-write expression (2) above to express it as follows:

$$E(m_{ijt}) = \frac{e^{w_{ijt}-c_{ijt}}}{\sum_{l \in D} e^{w_{ilt}-c_{ilt}}} s_{it}.$$ (3)

RUM models usually assume that the deterministic component of utility does not change with the origin country *i*. This allows us to re-write equation (3) as:

$$E(m_{ijt}) = \Phi_{ijt} \frac{y_{jt}}{\Omega_{it}} s_{it},$$ (4)

where $\Phi_{ijt} = e^{-c_{ijt}}$, $y_{jt} = e^{w_{jt}}$, and $\Omega_{it} = \sum_{l \in D} \Phi_{ilt} y_{lt}$. In this expression, migration depends on the accessibility Φ_{ijt} of destination *j*, its attractiveness y_{jt}, the capacity the origin country *i* has to send out migrants, proxied by its total population, s_{it}, and is inversely related to the utility derived by migrating to other destinations l $\in$ D or staying in the home country, Ω_{it}. Expression (4) is similar to other canonical gravity specifications, such as that used in the context of trade in Baier et al. (2019).

4.2 Main migration-gravity econometric specification

As is commonly done in the literature, we use GDP per capita levels (at PPP) as our measure of development levels at origin. To compute it, we use expenditure-side national GDP, which is most suitable for comparing living standards over time and across countries (Feenstra *et al.*, 2015), divided by total population size. We include both a linear and squared origin country GDP per capita variable in order to test for the hypothesized nonlinearity in the impact of development at origin on subsequent emigration flows. These are our two variables of interest.

Some econometric studies, however, claim that using merely a squared term in order to test for an (inverted) U-shaped relationship might lead to false conclusions (Lind and Mehlum, 2010; Haans *et al.*, 2016). Therefore, before we conclude that there truly is a U-shaped relationship, we consider the three-step procedure of Lind and Mehlum (2010) and test our model fit when including a cubic term to the empirical specification, as suggested in Haans *et al.* (2016).

To conform with the theory behind the RUM model (equation 4), we also control for population size. Within the RUM framework, population size measures the capacity that a given origin country has to send out migrants. Naturally, when a country has a larger population, it also has potentially higher migration flows in absolute numbers.

Following Rojas-Romagosa and Bollen (2018), we include country-pair fixed effects (FE) in our estimation. This is needed in order to account for all observable or unobservable bilateral time-invariant migration cost components, such as cultural or geographical distance, or any other time-invariant factor that might affect one's choice of destination j.

Taking logs of the RUM expression (4) above yields the following econometric specification:

$$ln(m_{ijt}) = \beta_1 ln(GDPpc_{it-10}) + \beta_2 [ln(GDPpc_{i,t-10})]^2 + \beta_4 ln(s_{it}) + I_{ij} + I_{jt} + I_i + \varepsilon_{ijt} \tag{5}$$

where m_{ijt} represents migration flows from country i to country j at time t; $GDPpc_{it-10}$ is the 10-year lag of GDP per capita at origin; s_{it} is the population size at origin at time t; I_{ij}, I_{jt} and I_i are respectively pair, destination-time and origin FE; ϵ_{ijt} is the error term.

Without taking logs as in (5) the empirical specification would run the risk of suffering from biased estimates due to the large number of zeros in our data set. Given the logarithmic form of our dependent variable, all pairwise observations with zero migration in the data would normally get dropped, as in log-linearized models estimated using OLS (e.g. Ortega and Peri, 2013; Llull, 2016). In order to avoid this, we estimate specification (5) using a Poisson pseudo-maximum-likelihood with high-dimensional fixed-effects (PPML-HDFE) estimator. As shown in Silva and Tenreyro (2006, 2011), PPML estimations perform well even when the proportion of zeros in the dependent variable is high. This justifies this approach given our data set. When compared to log-linearized gravity models, PPML estimations also yield relatively small biases in the presence of heteroscedasticity.

To estimate the above model, we employ the estimator by Correia *et al.* (2019). This estimator allows for a large set of different high-dimensional fixed effects structures. Exponentiating expression (5), our PPML migration specification can be expressed as follows:

$$m_{ijt} = \exp\{\beta_1 ln(GDPpc_{it-10}) + \beta_2 [ln(GDPpc_{i,t-10})]^2 + \beta_4 ln(s_{it})\} + I_{ij} + I_{jt} + I_i + \varepsilon_{ijt} \tag{6}$$

We use robust heteroscedasticity and autocorrelation consistent (HAC) standard errors and we cluster these around countries of origin. This is because our standard errors may be heteroscedastic and are probably correlated over time within origin countries' observations.

4.3 Dealing with endogeneity

A serious issue in the literature concerns the potential endogeneity. In particular, the possible reverse causality between development at origin and migration flows. The RUM expression (3) above does not make any specific assumptions about the direction of causality of the relationship between the prospective net utility of moving, $w_{ijt} - c_{ijt}$, and expected migration flows $E(m_{ijt})$. The former can impact the latter, but the reverse may also plausibly hold. Development at origin might affect one's migration aspirations and capabilities, and thus overall migration flows, through the channels mentioned in section 2. However, migration outflows can also affect development levels at origin. This could either happen directly (through remittances, modifications in consumption patterns, changes in asset accumulation at home, and

brain drain) or indirectly (for instance, through changes in the prices of local production factors and goods, or thanks to migrants encouraging investments into their areas of origin).[16]

One way in which the literature (imperfectly) accounts for endogeneity is by assuming that current migration outflows may only affect present and future development levels, while past levels of income per capita can affect future levels of emigration (Mayda, 2010; Ortega and Peri, 2013; Idu, 2019). That is, migration flows in year t, m_{ijt}, can only impact GDP per capita at t, $t + 1$, $t + 2$, …, while income in previous periods $t - 1$, $t - 2$, … may impact contemporaneous and future migration flows. Following the literature, we therefore relate current migration flows to lagged values of GDP per capita in our estimations. This reverse causality problem is likely to be less present in our case, as we use 10-year lags in GDP per capita.[17]

Another potential concern is the so-called multilateral resistance to migration (MRM). This is defined in Bertoli and Fernández-Huertas Moraga (2013) as the confounding influence that all potential alternative destinations $l \in D$ might have on one's choice to migrate to country j. This is encapsulated in the term Ω_{it} in equation (4). Ignoring this 'third country effect' has been shown to lead to omitted variable bias (Bertoli and Fernández-Huertas Moraga, 2013).

Existing strategies used in the literature to control for MRM do not work in this case. For example, Ortega and Peri (2013) control for heterogeneous preferences for migration across countries, which induce MRM by employing origin-time fixed effects. These are nonetheless perfectly collinear with any vector of time-varying origin variables w_{it} and therefore do not allow for the inclusion of development at origin, our variable of interest, into the model. A more general and less restrictive approach is the common correlated effects (CCE) estimator. This allows for consistent estimations in the case of spatially and serially correlated error structures. This estimator was proposed by Pesaran (2006) and employed in Bertoli and Fernandez-Huertas Moraga (2013). However, with only six time periods, our data set does not have a sufficiently large longitudinal dimension for the CCE estimator to be used here.

Following Mayda's (2010) approach and the arguments put forth in Beine *et al.* (2016), we (partially) control for MRM by introducing origin and destination-time fixed-effects. These absorb time-invariant and time-varying unobserved country-specific effects, respectively. They also serve as a proxy for MRM induced by time-invariant aspects of heterogeneous preferences for migration at origin or by the temporally fluctuating attractiveness of alternative destinations (Beine and Parsons, 2015). Origin FE are not collinear with GDP per capita at origin, which varies temporally, and can thus be included in the estimation model. This is analogous to the standard Anderson-Van Wincoop trade-gravity specification (2003), which incorporates importer and exporter fixed effects to account for multilateral resistance to trade.

Adequately accounting for MRM would require including origin-time fixed effects in our model along with destination-time and country-pair-varying fixed effects, as is done in state-of-the-art trade-gravity specifications, such as Baier *et al.* (2019). However, this would cause collinearity issues with respect to our variable of interest, which varies in both country and time dimensions. For this reason, we cannot fully

---

[16] See Mendola (2012) for a review of this literature.

[17] GDP per capita is averaged over decades, as explained in section 2. This means that we use information on GDP per capita from $t - 20$ to $t - 11$ to compute GDP per capita at $t - 10$. This longer time lag reduces the probability that reverse causality might be an issue in this case.

account for MRM and thereby eliminate the endogeneity bias from our estimation. Accordingly, our findings regarding to the migration-development nexus cannot be argued to represent a causal relationship.

5 Results

5.1 Main results

Table 1 shows the results from our main specification. The significant coefficients on the linear and squared GDP per capita terms have a positive and a negative sign, respectively, see column (2). These results provide empirical evidence of an inverted U-shaped relationship between GDP per capita at origin and emigration flows.

Moreover, it confirms the existence of the hump not only in the case of South-North flows, which had largely been the focus of past research on the topic, but for all combinations of origin and destination countries. By focusing on South-North flows, usually by leaving out non-OECD destinations from their analysis, previous studies have excluded about half of total international migration over the 1970-2020 period. By including such flows, we can therefore provide a more accurate test of the migration transition hypothesis, which is expected to hold for every origin globally.

The results from our model estimation on alternative time subsamples suggest that the migration hump holds both before and after 2000. Table 1 (columns (3) and (4)) shows the results for both the 1970-2000 period and the 2000-2020 period. As can be seen in the table, the coefficients on the linear and squared GDP per capita term are again significant and have a positive and negative sign, respectively. While the size of the two coefficients is lower for the latter timeframe, this decline is not significant.

Table 1   Results from base model, full sample and time subsamples

Migration flow	(1)	(2)	(3)	(4)
Ln GDPpc orig. ($t-1$)	-0.0661	4.033***	4.003**	3.366***
	(0.140)	(0.862)	(1.778)	(1.217)
Ln GDPpc orig. sq. ($t-1$)		-0.257***	-0.253**	-0.195***
		(0.0489)	(0.106)	(0.0669)
Ln pop. orig.	0.555*	-0.0588	0.827	-0.238
	(0.299)	(0.272)	(0.551)	(0.402)
Year sample	1960-2020	1960-2020	1960-2000	2000-2020
Observations	89,490	89,490	65,703	30,812
Pseudo R-squared	0.899	0.902	0.927	0.907

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Destination-time, country pair and origin fixed effects are included in all estimations.

The finding of a migration hump remains robust when estimating the model separately for each decade within the 1970-2020 timeframe. This is illustrated in Figure 3, which shows the results of our non-parametric cross-country regressions of emigrant stocks on GDP per capita at origin (PPP-adjusted), for each of the five decades within the 1970-2020 period. Emigration rates are computed as the ratio of the total number (stock) of migrants from a given country residing in a foreign country to the total population

in the origin country. These regressions depict an inverted U-shaped relationship between development at origin and emigration for each of these decades. Our results confirm those found in Clemens (2014) and Dao *et al.* (2018).

Figure 3　　　Non-parametric regression of the migration-development nexus in cross-section for each year in the 1970-2020 time period

Note: The dark red lines depict Second-Order Gaussian continuous kernel non-parametric regressions. Countries with emigration rates that are higher than 1 per year are omitted. The Cayman Islands and Kuwait are omitted from the regressions as well.

5.2　Robustness analyses

In order to test the robustness of our results, we perform two sets of robustness analyses. First, in Section 5.2.1, we test the robustness of the hump shape as a whole by conducting the analysis with several alternative specifications, which all support the finding of the hump shape. Second, in Section 5.2.2, we specifically test the robustness of the finding that emigration initially rises when a low-income country begins to develop, corresponding to the upward sloping 'left hand side' of the migration hump at the lower end of the income distribution. There, we do not find support for the initial increase of emigration with development.

5.2.1　Robustness analyses of the hump shape

To test the robustness of the hump shape we use several alternative specifications. These include: (i) the addition of several origin-time control variables (their definitions and sources can be found in the Appendix Table A. 2, along with GDP per capita and population at origin), in order to prevent omitted

variable bias in the origin country-time dimension; (ii) the inclusion of an interaction term between geographical distance and income at origin, and (iii) additional tests of the existence of an inverted U-shape between GDP per capita at origin and emigration flows.

(a) Controlling for demographic and other origin-time variables

In the first set of robustness checks, we augment our base model with several socio-demographic control variables. These serve to enrich our model by capturing more of the variation in the origin-time dimension and effectively reduce potential omitted variable bias issues.

First, demographic factors at origin can significantly influence migration patterns through their impact on the domestic labor market structure. On a global scale, inter-country differentials in demographic structures might affect the directionality of migration flows, whereby countries with a large inactive population demand more labor from abroad in order to support the economy, while residents of countries with a relatively large labor force are more willing to emigrate. Also, higher population densities can make one more willing to emigrate, as it limits the amount of available resources per person. In this light, we introduce the age dependency ratio and population density at origin (both defined in Appendix Table A. 2) as controls. We expect a positive sign on the coefficient on population density: an increase entails higher pressures on a country's resources, potentially leading to higher rates of emigration. The coefficient for age dependency could be both positive (e.g. a higher elderly dependency could lead to more emigration among pensioners, while a higher youth dependency could lead to more pressure for parents to look for better income opportunities abroad) or negative (e.g. higher elderly dependency may require more immigrants in elderly care), since several mechanisms are at play here.

Moreover, political instability or poor governance may catalyze emigration, sometimes by forcing it. The landscape of politically driven emigration can range from people fleeing a war or a genocide to those seeking better living conditions, in the form of secured property rights or the freedom of expression. In order to capture the influence of these factors on emigration, we introduce the Polity IV index at origin, along with the number of months the origin country has been in any sort of conflict (genocides, politicides, and ethnic and revolutionary wars). We expect a negative sign for the former, as one's willingness to migrate in a relatively democratic country is expected to be low. The coefficient on our conflict variable is expected to be positive.

Populations can be displaced by natural disasters as well, which might destroy means of living in the origin country and thereby force people to flee appalling conditions at home to seek higher material wealth abroad. We account for these in an alternative specification through the number of natural disasters that occurred in a country during the time period considered. The coefficient on this variable is expected to be positive, as the rise in natural disaster occurrences in a given time period should lead to more outward migration.

In order to prevent potential collinearity issues, only control variables that have an absolute correlation of 0.4 or lower with GDP per capita are included in the estimation.[18] Further, since natural disaster occurrences are highly correlated with the natural logarithm of population at origin (correlation > 0.4), we do not simultaneously include them in the estimation. The same goes for the Polity IV index at

---

[18] A correlation matrix for all explanatory variables included can be found in Appendix Table A. 3.

origin and the age dependency ratio. Therefore, we first incorporate each control variable to the main model separately in order to test their significance with no influence from other potential factors. We then include all controls at the same time, excluding some variables to avoid collinearity.[19]

(b) Estimating alternative time subsamples

The on average positive global growth rates in GDP per capita between 1970-2020 led to a rightward shift of the world per capita income distribution. An increasing number of countries now lie in the middle- to high-income per capita group. This can have an impact on the existence of the migration transition. If the migration turning point lies at relatively low GDP per capita levels, then the hump will be more pronounced for earlier periods, assuming that the turning point remains constant over time. Otherwise, if the turning point does move to the right over time, this effect does not occur.

In order to test whether the hump shape became less pronounced, we subdivide our country-time sample into two distinct timeframes, taking advantage of the panel structure of our data set. The two timeframes chosen were 1970-2000 and 2000-2020 (the year 2000 cutoff was chosen arbitrarily). We then estimate model (6) on these two subsamples. This will also enable us to have a better idea of where the actual migration transition point lies, and thus which income levels actually drive the migration transition.

(c) Controlling for interactions between geographical distance and income at origin

Furthermore, the impact of a change in income at home on emigration might be different depending on the distance to potential migration destinations chosen by a would-be migrant. For instance, the effect of a positive income shock on one's decision to move might be more pronounced if the destination considered is closer to home. This can be due to the fact that migrants considering a faraway destination might focus more on long-run income prospects than fluctuating income shocks in their migration decision. Moving farther away implies less flexibility to move back and forth to one's home country to benefit from wage fluctuations. Following Lull (2016), the interaction between geographic distance and income at origin, $\widehat{GDPpc_{it}}.\widehat{D_{ij}}$, is included into model (6), where, for any $\bar{x}$ being the sample mean of variable x, we define $\hat{x} \equiv x - \bar{x}$. This yields the following estimation model:

$$m_{ijt} = \exp\{\beta_1 \ln(GDPpc_{it-1}) + \beta_2 [\ln (GDPpc_{i,t-1})]^2 + \beta_3 \ln (s_{it}) + \beta_5 \widehat{GDPpc_{it}}.\widehat{D_{ij}}\}$$
$$+ I_{jt} + I_i + \epsilon_{ijt} \tag{7}$$

(d) Additional tests for the existence of an inverted U-shape

Lastly, we conduct further statistical tests of the inverted U-shaped relationship between development at origin and emigration. As argued in Lind and Mehlum (2010), merely adding a quadratic term to an otherwise linear specification can be too weak a criterion to test for such a nonlinear relationship if the latter is either convex or monotone. In this case, one might be led to a type I error where the null hypothesis of linearity is wrongly rejected because an extreme point is found and thus an inverted U-shape.

---

[19] The Polity IV index or age dependency (correlation coefficient of -0.44) and the natural disaster variable or natural logarithm of total population (correlation coefficient of 0.62) are excluded in turn because of their relatively high correlation with each other.

To account for this potential issue, we follow Lind and Mehlum's (2010) three-step procedure.[20] First, we verify that β_2 in specification (6) is significantly negative. Second, we check whether the slopes at both ends of the data range, to the right and to the left of the optimum, are significantly different from zero, and positive and negative, respectively. Third, the turning point should lie well within the data range.

With regard to the first step, we use the results from the estimation of the empirical specification (6). The second and third step are done using the Sasabuchi test (Sasabuchi, 1980; Lind and Melhum, 2010). This test checks the robustness of an inverted U-shaped relationship by testing whether the slopes to the left and the right of the turning point are significantly positive and negative, respectively. We also choose to test the fit of our model when adding a cubic term, thus allowing for the curve to take an S-shape rather than a U-shape.

Results of robustness checks

The results of the first robustness checks are that the main result remains unchanged when augmenting the main model with a set of socio-demographic controls. As shown in Table 2, in terms of significance and sign, our main result regarding the two GDP per capita coefficients remains unchanged when the age dependency ratio, the Polity IV index, the number of natural disaster occurrences, conflict duration and population density are individually added to the main model.

Networks and population density, which are the only variables with significant coefficients, both have the expected positive sign. This suggests that these variables, either through an increase of a country's diaspora population or through a negative impact on resource availability, might foster emigration.

Table 2   Results from the estimation of the base model, augmented with selected origin-time control variables (specification is sometimes changed to avoid multicollinearity)

Migration flow	(1)	(2)	(3)	(4)	(5)	(6)
Ln GDPpc ($t-1$)	4.090***	4.387***	3.855***	3.308***	3.964***	4.409***
	(0.864)	(0.911)	(0.827)	(0.925)	(0.822)	(0.861)
Ln GDPpc sq. ($t-1$)	-0.260***	-0.278***	-0.246***	-0.209***	-0.252***	-0.281***
	(0.0487)	(0.0521)	(0.0464)	(0.0524)	(0.0467)	(0.0490)
Ln population	-0.0357	-0.172		-0.349	-0.0652	-0.170
	(0.273)	(0.263)		(0.313)	(0.277)	(0.275)
Age dependency ratio	0.0018					
	(0.0052)					
Polity IV index		0.00230				
		(0.0115)				
Nat. disaster occurrence			-0.0010			
			(0.0015)			
Networks ($t-2$)				0.409**		
				(0.165?)		
Conflict duration					0.0026	

---

[20] See also Haans *et al.* (2016).

						(0.0018)
Population density						0.0004***
						(0.0001)
Observations	88,229	80,692	90,396	75,123	89,490	88,171
Pseudo R-squared	0.902	0.903	0.902	0.907	0.902	0.902

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Destination-time, country pair and origin fixed effects are included in all estimations.

Including all of these control variables along with the main model, changing the specification to avoid multicollinearity issues[21] shows that the main results do not change. As Table 3 shows, both the linear and squared GDP per capita terms remain highly significantly positive and negative, respectively. Moreover, population density and our network variable are both significant across most specifications. This evidences the role of demographic pressure and the impact of the size of the diaspora in affecting one's propensity to migrate. Population density keeps its expected positive sign, while the coefficient of the network variable on occasion however unexpectedly turns negative.

Table 3   Results from the estimation of the base model with different combinations of control variables

Migration flow	(1)	(2)	(3)	(4)	(5)
Ln GDPpc ($t-1$)	3.653***	3.423***	3.304***	2.986***	3.221***
	(0.879)	(0.900)	(0.839)	(0.902)	(0.844)
Ln GDPpc sqr. ($t-1$)	-0.230***	-0.219***	-0.206***	-0.188***	-0.207***
	(0.050)	(0.053)	(0.047)	(0.052)	(0.050)
Ln population	-0.403	-0.555*			-0.613**
	(0.328)	(0.302)			(0.301)
Age dependency ratio	0.00130		0.00284		-0.00418
	(0.00611)		(0.00628)		(0.00616)
Polity IV index		0.0134		0.0143	0.0136
		(0.0109)		(0.0114)	(0.0108)
Conflict duration	0.00256	0.00267	0.00250	0.00270	0.00264
	(0.00195)	(0.00178)	(0.00204)	(0.00183)	(0.00176)
Network ($t-2$)	0.422***	-16.950***	0.434***	-16.810***	-17.330***
	(0.155)	(5.071)	(0.149)	(5.196)	(5.333)
Population density	0.000321***	0.000415**	0.000304**	0.000291	0.000394**
	(0.000117)	(0.000201)	(0.000125)	(0.000207)	(0.000183)
Natural disaster occurrences			0.00167	0.000296	6.04e-05
			(0.00226)	(0.00214)	(0.00202)

[21] Variables with more than 0.4 (absolute) correlation are not included together in the same specification. A correlation matrix can be found in Table A.3 in the Appendix.

Observations	72,775	66,973	72,775	66,973	66,973
Pseudo R-squared	0.907	0.911	0.907	0.911	0.911

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Destination-time, country pair and origin fixed effects are included in all estimations.

The results from the estimation of model (7) are shown in Table 4. While both the linear and the squared GDP per capita terms are highly significant and have the expected positive and negative signs, respectively, the added interaction term is weakly significant. In accordance with Lull (2016), we therefore find some evidence suggesting that income shocks might have a heterogeneous impact on emigration depending on distance to destination.

The results of the Sasabuchi test for the (inverted) U-shape can be found in the Appendix Table A. 5. The slopes at both ends of the data ranges are significant, and of the expected signs: positive at the lower bound and negative at the upper bound. The overall test for the presence of an inverted U-shape between GDP per capita and migration flows also enables us to reject the null hypothesis that emigration evolves linearly with GDP per capita, and thus further confirms the existence of the inverted-U shape. Moreover, the extremum point, at $\ln(\text{GDPpc}_{it}) = 7.85788$, lies well within the Ln GDP per capita range, which goes from 6.126 to 12.541 (see Appendix Table A. 4 for summary statistics). Finally, adding a cubic term to the empirical model does not improve model fit, as Table A. 7 in the Appendix depicts. The linear, squared and cubed GDP per capita terms are insignificant. Given these results and the ones above, the migration-development nexus is thus more likely to follow an inverted-U shape than an S-shape.

Table 4   Results from the estimation of model (7) with the interaction term distance and GDP per capita

	Migration flow
Ln GDPpc orig. $(t-1)$	3.913***
	(0.910)
Ln GDPpc orig. sq. $(t-1)$	-0.248***
	(0.052)
Ln pop. orig.	-0.103
	(0.291)
Ln Dist.*Ln GDPpc orig.	0.136*
	(0.077)
Observations	84,336
Pseudo R-squared	0.904

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Destination-time, country pair and origin fixed effects are included in all estimations.

5.2.2 Robustness analyses of the initial increase of migration with development

All the findings of our robustness tests in sections 5.1 and 5.2.1 seem to suggest that there is strong empirical support for the migration transition hypothesis' prediction of a migration hump: an inverted-U relationship between development levels and emigration. This finding is consistent with our Figures 2 and 3 which show that middle-income countries tend to have higher emigration rates than either low-income or high-income countries. Recently, several authors (e.g. De Haas (2019), Clemens and Postel (2018)) have concluded from this finding of such an inverted-U relation that this implies that, as low-income countries develop, their emigration will tend to increase first before declining only after some threshold level of income.

If this conclusion holds for individual countries, then this could have serious implications for development programs. In particular, it would imply that development cooperation, to the extent that it contributes to economic development, contributes to increased emigration from low-income to high-income countries. As the authors mentioned above have pointed out, development cooperation in that case is not a proper instrument to reduce emigration from low-income countries.

However, even if the migration hump finding is as robust as it seems to be, can we actually conclude that all individual countries will follow this inverted U-pattern as they grow richer? In other words, will emigration for an individual low-income country indeed rise as it starts developing economically, and fall after some threshold middle-income level? The answer is that this does not necessarily follow from the finding of an inverted-U relation between development and emigration based on cross-country or panel data. Benček and Schneiderheinze (2019) are therefore critical of any causal interpretation of the migration hump.

One reason why the cross-sectional evidence for the hump shape does not necessarily demonstrate an individual country's transition path is that, while middle-income countries experience higher emigration than low-income (and high-income) countries, this is not necessarily due to their income differences. It may also be due to fundamental heterogeneity between the different country income groups that simultaneously affect both economic development and migration (Lucas, 2019). If such omitted variables are driving the inverted U-relationship, then the migration hump is misinterpreted as being a result of economic development.

This point is not solely relevant for evidence of an inverted-U relationship based on cross-section data, but also for evidence of a migration hump based on panel data, as we are using in this paper. The reason is as follows. By using panel data, we exploit both the variation over time and the variation across countries. The variation over time for each country across the income distribution is however limited in the sense that even though we use a large 50-year timeframe from 1970 to 2020, there is no country that has covered the whole income distribution over this period developing from a low-income to a high-income country. Despite substantial economic growth for many countries within this period, countries have still moved within a limited range of the income distribution. This implies that, even though we are using panel data and exploiting some income variation over time for each country, we are still to a large extent relying on the cross-section variation in the data for our finding of an inverted-U relation between emigration and economic development. That means that this finding of the inverted-U relation is to an important extent still driven by the fact that middle-income countries experience higher emigration than low- or high-income countries. So again, the conclusion that income levels are driving the inverted-U relation between

development and migration will not necessarily hold if there is systematic heterogeneity across countries in these income groups. This is particularly the case if there is heterogeneity with respect to factors that affect both development and migration, and if these factors are not properly controlled for.

One reason why full control for all relevant factors affecting both income and emigration is complicated in all panel data studies on emigration and development is the following. Even though in our above panel data analysis we have applied a very extensive set of control variables, including several origin-time control variables and destination-time, country-pair, and origin fixed effects, there might still be some origin-time factors that affect both emigration and development and hence require additional controls. While such factors in principle could be controlled for by using origin-time fixed effects, such fixed effects are however perfectly collinear with any origin-time varying variables and hence cannot be simultaneously included in the specification with development at origin, which is our variable of interest. As indicated, this issue is relevant for all panel data studies on emigration and development. Therefore, additional robustness checks are required in order to test whether low-income countries as they develop indeed initially experience an increase in emigration due to economic growth. This will be done in the next section.

In order to avoid this issue of inappropriately using the higher emigration levels of middle-income countries compared to low-income countries, while not being able to fully control for fundamental differences between the two groups of countries that may drive the result of an initial increase in emigration with development, we perform several robustness tests in this subsection that all relate to the upward-sloping part of the migration hump in order to test whether as low-income countries grow, their emigration will tend to initially increase.

The first test we perform is for the subsample of 46 countries that actually transitioned from low-income to middle-income status. We test whether emigration from these countries increased with development, by applying our base regression model on this sub-group only. In this case, the included middle-income countries are the same as the included low-income countries (only at a later point in time) and hence there is no heterogeneity between the two income groups when using this subsample. This subsample consists of 46 countries that have all developed from the low-income to the middle-income category in the period 1970-2020 according to the World Bank income classification.[22] If emigration initially increases with economic development until low-income countries reach some middle-income threshold level, then we would expect a positive and significant coefficient on our linear GDP per capita variable. We perform the regression both with and without a squared term for our GDP per capita variable. The results from our estimation for this subsample of countries that have transitioned from low-income to middle-income are shown in columns (1) and (2) of Table 5. The table shows that in neither case, we get a significantly positive coefficient for our GDP per capita variable and hence we cannot conclude that for this group of countries economic development has resulted in an increase of emigration. We have also

---

[22] The countries included in this subsample are Angola, Albania, Armenia, Azerbaijan, Bangladesh, Bhutan, Cambodia, Cameroon, China, the Comoros, Republic of Congo, Côte d'Ivoire, the Arab Republic of Egypt, Equatorial Guinea, Georgia, Ghana, Guyana, Honduras, Indonesia, India, Kenya, Kyrgyzstan, Lao PDR, Lesotho, Maldives, Mauritania, Republic of Moldova, Mongolia, Myanmar, Nicaragua, Nigeria, Pakistan, Papua New Guinea, São Tomé and Príncipe, Senegal, Solomon Islands, Sri Lanka, Sudan, Tajikistan, Timor-Leste, Turkmenistan, Ukraine, Uzbekistan, Vietnam, the Republic of Yemen, Zambia.

performed the regression for several subsets of this group of 46 countries and the results are all similar in the sense that they show no evidence of an increase of emigration with development for these countries.

Next, we perform the same test for the similar sample of countries that have transitioned from the low-income to the middle-income category, but now excluding China and India. These two countries are outliers in terms of population and country size, which may have important implications for emigration, and they have also experienced relatively high economic growth. The results for this subsample are presented in columns (3) and (4) of Table 5. The results in column (4) for the regression including the quadratic term for our GDP per capita variable show no significance for the coefficients on either the linear or squared GDP per capita variable. However, the results in column (3) for the regression including only the linear GDP per capita variable show a very significant and negative coefficient on our GDP per capita variable.

Limiting our analysis to this sample of 44 countries that have actually developed from being a low-income country to becoming a middle-income country, the finding is thus that emigration has not increased but rather declined with economic development. By focusing solely on the countries that actually made the transition from low-income to middle-income status, we avoid the issue of inappropriately using the higher emigration levels of middle-income countries compared to low-income countries, while not being able to fully control for fundamental differences between the two groups of countries. For this relevant subsample, it is clear that, when low-income countries develop economically, their emigration declines. This obviously has important policy implications as it refutes the recent belief that development programs contribute to rising emigration when promoting economic development.

In addition to this subsample of countries that each developed from low-income to middle-income status, we also test whether there is an increase in emigration with development for the subsample of all African countries. This is also an interesting subsample because these countries have grown in the covered 50-year period from being mostly low-income to being mostly lower-middle income, with less than half of the countries still being low-income countries in 2020 and a few countries transitioning to the upper-middle income category. Mean GDP per capita for African countries increased substantially from US$ 1,738 to US$ 4,798 during this period.[23]

The results for our subsample of African origin countries confirm that GDP per capita growth does not give rise to emigration from African countries. The results are presented in columns (5) and (6) of Table 5. The results show that, despite substantial increase in GDP per capita among African countries, there is no sign of a significant positive relation between GDP per capita and emigration, as some authors in the migration literature suggested. Instead, the relationship is negative though not significant. As shown in the table, population at origin does show a significant and positive coefficient. This indicates that population growth may have been driving higher emigration levels for African countries.

Table 5   Results from base model, countries that transitioned from LIC to MIC and African countries

Migration flow	(1)	(2)	(3)	(4)	(5)	(6)
Ln GDPpc orig. ($t-1$)	-0.253	0.245	-0.470***	0.230	-0.086	-0.339

[23] In constant 2011 U.S. dollars.

	(0.215)	(2.029)	(0.162)	(2.263)	(0.112)	(0.774)
Ln GDPpc orig. sq. $(t-1)$		-0.033		-0.0461		0.017
		(0.129)		(0.146)		(0.052)
Ln pop. orig.	-0.505	-0.522	-0.239	-0.232	1.290***	1.309***
	(0.370)	(0.391)	(0.444)	(0.451)	(0.417)	(0.419)
Subsample	LIC to MIC	LIC to MIC	LIC to MIC excl China, India	LIC to MIC excl China, India	Africa	Africa
Observations	18,524	18,524	16,423	16,423	22,012	22,012
Pseudo R-squared	0.935	0.935	0.934	0.934	0.915	0.915

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Destination-time, country pair and origin fixed effects are included in all estimations.

The next robustness check is to test whether, for the lower part of the income distribution, there is an upward-sloping 'left hand side' of the migration hump, in other words, whether there is a significantly positive relation between development and emigration up to a certain point. We first check this for our base model applied to the full sample of countries for which the result was presented in column (2) in Table 1. The corresponding extreme point for this base model result lies at a per capita GDP of US$ 2,586. We therefore now test our migration base model (both only with a linear and also with a quadratic term for GDP per capita) applied on all observations in our data set up to this extreme point for GDP per capita. The results are presented in column (1) and (2) of Table 6 and do not show a significantly positive coefficient for our linear GDP per capita term that we would expect if an increase in income would lead to more emigration in this lower part of the income distribution until the extreme point of the hump. The coefficient on squared GDP per capita is also insignificant.

We perform a similar test for the highest extreme point of GDP per capita that we found across all other specifications used in sections 4.2 and 4.3.1 and that is the one applied for the time subsample 2000-2020, for which the results were shown in column (4) of Table 1. The extreme point corresponding to this result lies at a GDP per capita of US$ 5,693. We again test our base model of emigration, again both with only a linear and also with a quadratic term for GDP per capita, applied on all observations below this turning point of the hump of US$ 5,693. The results are presented in columns (3) and (4) of Table 6 below and show that also using this extreme point, the coefficients on our GDP per capita variable are insignificant, indicating there is no significant relation between income per capita and emigration at this part of the income distribution. We also tested the extreme points for all other specifications used in section 4.3.1 and since these all lie to the left of the above extreme point of US$ 5,693, the results from the estimation of our base emigration model applied to the observations to the left of these respective extreme points are all similar in the sense that they do not show a positive and significant relation between our GDP per capita variable and emigration.

Table 6   Results from the estimation of the base model, up to various extreme points found

Migration flow	(1)	(2)	(3)	(4)
Ln GDPpc $(t-1)$	-0.064	0.389	-0.137	1.612

	(0.158)	(4.028)	(0.142)	(2.206)
Ln GDPpc sq. $(t-1)$		-0.032		-0.117
		(0.282)		(0.146)
Ln population	-0.122	-0.113	-1.194***	-1.213***
	(0.483)	(0.468)	(0.448)	(0.446)
Subsample	GDP pc extreme point of US$2586	GDP pc extreme point of US$2586	GDP pc extreme point of US$5693	GDP pc extreme point of US$5693
Observations	25,864	25,864	45,909	45,909
Pseudo R-squared	0.952	0.952	0.937	0.937

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Destination-time, country pair and origin fixed effects are included in all estimations.

Next, we test our base model on all observations in the first and second quartile of the income distribution. Again, if emigration initially increases with development, we would expect to find a significant and positive coefficient on our GDP per capita variable. The results can be found in columns (1) and (2) of Table 7 and show no significance for either the linear or the squared GDP per capita variable.

Finally, we apply the PPML base model on all observations with a maximum GDP per capita of US$ 9,999, which happens to be the mean of GDP per capita across all upper middle-income countries. Columns (3) and (4) of Table 7 show the results and indicate no significance for either the GDP per capita variable or for the hump-shape of the relation between development and emigration.

Table 7   Results from the estimation of the base model, up to various income thresholds

Migration flow	(1)	(2)	(3)	(4)
Ln GDPpc $(t-1)$	-0.020	-2.356	-0.087	2.214
	(0.143)	(2.221)	(0.144)	(1.512)
Ln GDPpc sq. $(t-1)$		0.157		-0.151
		(0.149)		(0.098)
Ln population	-1.210**	-1.308**	-0.702	-0.681
	(0.536)	(0.555)	(0.451)	(0.447)
Subsample	First and second quartile of income distribution	First and second quartile of income distribution	Until GDP pc of US$9999, mean income UM countries	Until GDP pc of US$9999, mean income UM countries
Observations	37,508	37,508	53,780	53,780
Pseudo R-squared	0.946	0.946	0.930	0.930

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Destination-time, country pair and origin fixed effects are included in all estimations.

The conclusion from these robustness checks is twofold. On the one hand, the finding of a hump-shaped relationship between emigration and development levels is highly robust in panel data settings, using data for 180 countries and a 50-year timeframe. On the other hand, it is not correct to conclude that, in any

given country, emigration initially increases with economic development before it starts to fall. In particular, the 'left hand side' of the migration hump does not withstand any of the robustness tests that we performed. On the contrary, when we focus on low-income countries that actually transitioned to middle-income status, we find evidence that emigration actually declined with economic development. This suggests that the inverted U-shaped relationship of economic development and migration cannot be interpreted as a causal relationship.

6 Concluding remarks

This paper has rigorously tested the migration transition hypothesis according to which emigration follows an inverted U-shaped relationship with economic development. The migration transition hypothesis suggests that emigration first increases, as countries move from low to middle-income levels of development, and subsequently decreases again as countries grow richer. As predicted by several migration transition theories, such a non-linear pattern could emerge from various factors at play, including financial constraints that diminish over time, migrant networks abroad that increase with migration, or a demographic transition.

In order to test this hypothesis, we applied a migration version of the gravity model, micro-founded by the Random Utility-Maximization (RUM) model, on a global panel data set comprising 180 origin and destination countries and a 50-year timeframe (1970-2020). This is the most extensive panel data set used so far in the literature to test for the existence of the migration hump. We used GDP per capita at origin as a proxy for development levels and include a linear and a squared term to account for the nonlinearities predicted by migration transition theories. We used the most recent PPML estimator and, following the literature, controlled for the influence of alternative destinations on one's decision to migrate (so-called multilateral resistance to migration). We did so by incorporating several origin-time control variables and various fixed effects structures controlling for unobserved origin-, destination-time, time and country-pair-characteristics potentially affecting migration flows.

Based on this panel data analysis, we find strong empirical support for an inverted-U relationship between emigration and development levels. Our results are robust to (a) the addition of several origin-time control variables, (b) the use of different time and country subsamples (with and without non-OECD countries), (c) the inclusion of an interaction term between geographical distance and income at origin and (d) several additional tests of the existence of an inverted-U shaped relation between GDP per capita at origin and emigration flow.

However, the finding of an inverted U-shaped relation between economic development and emigration is mainly driven by cross-country heterogeneity in factors other than income and therefore the migration hump cannot be interpreted as a causal relation. In several additional robustness analyses we found that, for a given low-income country, an increase in economic development does not lead to higher emigration. On the contrary, for a subsample of 44 countries that actually transitioned from low-income to middle-income status (excluding China and India), we even found evidence that emigration rather declined with economic development. Drawing the conclusion that the inverted-U relationship is causal therefore seems unfounded.

This new finding, supported by various robustness checks, has important policy implications. In contrast with what other authors (e.g. De Haas, 2019 and Clemens and Postel, 2018) have concluded based

on cross-sectional findings, we can no longer conclude that, as low-income countries develop, their emigration will tend to increase before declining after a certain middle-income turning point. While we do find empirical evidence of an inverted U-relation between economic development and emigration using the full sample of 180 countries over 50 years, it seems that this finding is driven by the underlying cross-sectional pattern of middle-income countries having higher emigration rates than either low- or high-income countries. These differences in emigration rates are likely caused by fundamental differences between countries in different income categories that make a causal inference of the inverted-U relation invalid.

Moreover, akin to other papers in the existing literature on this topic, we are not able to fully control for the potential endogeneity arising from the reversed causality between migration and GDP, nor from the multilateral resistance to migration, i.e. the unobserved impact of the attractiveness of alternative destinations on one's willingness to emigrate. Due to these issues any causal interpretation of the migration hump is unfounded. Although in our analysis we do not eliminate the bias due to endogeneity, we are able to reduce it by including a decade-to-decade lag of GDP per capita at origin as an instrument in order to tackle reverse causality and by using country-, destination-time and country-pair-varying fixed effects in order to partially account for multilateral resistance to migration.

We circumvent the remaining endogeneity problem due to fundamental differences between countries in different income categories that we cannot fully control for, by estimating the model solely for those countries that actually transitioned from low-income to middle-income status. In this case, the included middle-income countries are the same as the included low-income countries (only at a later point in time) and hence there is no heterogeneity between the two income groups when using this subsample. Interestingly, the results for this subsample (which excludes China and India) show importantly that emigration actually declines as low-income countries develop economically. This obviously has important policy implications: it suggests that development programs can in fact promote economic development in low-income countries without encouraging emigration.

References

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, *93*(1), 170–192.

Bade, J. and A. De Kemp (2019), *Migration and Development*, Dutch Ministry of Foreign Affairs

Baier, Scott L. & Yotov, Yoto V. & Zylkin, Thomas, 2019. On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration, *Journal of International Economics*, vol. 116(C), pages 206-226.

Beine, M., Bertoli, S., & Fernández-Huertas Moraga, J. (2016). A Practitioners' Guide to Gravity Models of International Migration. *The World Economy*, *39*(4), 496–512.

Beine, M., Boucher, A., Burgoon, B., Crock, M., Gest, J., Hiscox, M., … Thielemann, E. (2016). Comparing immigration policies: An overview from the IMPALA database. *International Migration Review*, *50*(4), 827–863.

Beine, M., Bourgeon, P., & Bricongne, J. (2019). Aggregate fluctuations and international migration. *The Scandinavian Journal of Economics*, *121*(1), 117–152.

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

Belot, M. V. K., & Hatton, T. J. (2012). Immigrant Selection in the OECD. *The Scandinavian Journal of Economics*, *114*(4), 1105–1128.

Benček, D., & Schneiderheinze, C. (2019). More development, less emigration to OECD countries: Identifying inconsistencies between cross-sectional and time-series estimates of the migration hump (No. 2145). *Kiel Working Paper.*

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

Borjas, G. (1987). Self-Selection and the Earnings of Immigrants. The American Economic Review, 77(4), 531-553. Retrieved from www.jstor.org/stable/1814529

Caselli, M. (2019). "Let Us Help Them at Home": Policies and Misunderstandings on Migrant Flows Across the Mediterranean Border. *Journal of International Migration and Integration*, 1–11.

Clemens, M. A. (2014). *Does Development Reduce Migration?* IZA Discussion Papers.

Clemens, M. A. & Postel, H.M. (2018). "Deterring Emigration with Foreign Aid: An Overview of Evidence from Low-Income Countries" *Population and Development Review*, *4*, 667-693.

Correia, S., Guimarães, P., & Zylkin, T. (2019). PPMLHDFE: Fast poisson estimation with high-dimensional fixed effects. *ArXiv Preprint ArXiv:1903.01690.*

Dao, T. H., Docquier, F., Parsons, C., & Peri, G. (2018). Migration and development: Dissecting the anatomy of the mobility transition. *Journal of Development Economics*, *132*, 88–101.

De Haas, H. (2007). Turning the tide? Why development will not stop migration. *Development and Change*, *38*(5), 819–841.

De Haas, H. (2010). *Migration Transitions. IMI Working Papers*. Oxford.

De Haas, H. (2011). *The Determinants of International Migration* (IMI Working Papers). Oxford.

De Haas, H. (2019). *Why Development Will Not Stop Migration.* https://www.macmillanihe.com/blog/post/why-development-will-not-stop-migration-hein-de-haas/.

De Haas, H., Natter, K., & Vezzoli, S. (2014). Compiling and coding migration policies: Insights from the DEMIG POLICY database. *International Migration Institute, DEMIG Project Paper*, *16*, 43.

Di Giovanni, J., Levchenko, A. A., & Ortega, F. (2015). A global view of cross-border migration. *Journal of the European Economic Association*, *13*(1), 168–202.

Djajic, S., Kirdar, M. G., & Vinogradova, A. (2016). Source-country earnings and emigration. *Journal of International Economics*, *99*, 46–67.

Egger, P., & Nigai, S. (2015). *Structural Gravity with Dummies Only.* CEPR Discussion Papers.

Epstein, G. S. (2008). Herd and network effects in migration decision-making. *Journal of Ethnic and Migration Studies*, *34*(4), 567–583.

European Commision. (2015). *A European Agenda on Migration.* Brussels.

Faini, R., & Venturini, A. (2010). Development and migration: Lessons from southern Europe. In *Frontiers of Economics and Globalization* (Vol. 8, pp. 105–136). Emerald Group Publishing Limited. https://doi.org/10.1108/S1574-8715(2010)0000008011

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, *105*(10), 3150–3182.

Flahaux, M.-L., & De Haas, H. (2016). African migration: trends, patterns, drivers. *Comparative Migration Studies*, *4*(1), 1.

Gonzalez-Garcia, M. J. R., Hitaj, M. E., Mlachila, M. M., Viseth, A., & Yenice, M. (2016). *Sub-Saharan African migration: patterns and spillovers.* International Monetary Fund.

Gould, J. D. (1979). European Inter-Continental Emigration 1815-1914: Patterns and Causes. *Journal of European Economic History*, *8*(3), 593–679.

Grogger, J., & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, *95*(1), 42–57.

Guha-Sapir, D. (2019). EM-DAT: The Emergency Events Database - Université Catholique de Louvain (UCL) - CRED.

Haans, R. F., Pieters, C., & He, Z. L. (2016). Thinking about U: Theorizing and testing U-and inverted U-shaped relationships in strategy research. *Strategic Management Journal*, *37*(7), 1177-1195.

Harris, N. (2002). Thinking the Unthinkable: The Immigration Myth Exposed (IB Tauris) London.

Hatton, T. J., & Williamson, J. G. (2011). Are third world emigration forces abating? *World Development*, *39*(1), 20–32.

Head, K., Mayer, T., & Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, *81*(1), 1–14.

Héran, F. (2018). Europe and the spectre of sub-Saharan migration. *Population & Sociétés*, (558).

Idu, R. (2019). Source Country Economic Development and Dynamics of the Skill Composition of Emigration. *Economies*, *7*(1), 18.

International Organization for Migration. (2011). *Glossary on Migration.* (R. Perruchoud & J. Redpath-Cross,

Eds.) (2nd ed.).

Khoudour-Castéras, D. (2009). Neither migration nor development: The contradictions of French co-development policy.

Larch, M., Wanner, J., Yotov, Y. V, & Zylkin, T. (2019). Currency Unions and Trade: A PPML Re-assessment with High-dimensional Fixed Effects. *Oxford Bulletin of Economics and Statistics*, *81*(3), 487–510.

Letouzé, E., Purser, M., Rodríguez, F., & Cummins, M. (2009). Revisiting the migration-development nexus: a gravity model approach. *Human Development Research Paper (HDRP) Series*, *44*.

Lind, J. T., & Mehlum, H. (2010). With or without U? The appropriate test for a U-shaped relationship. *Oxford bulletin of economics and statistics*, *72*(1), 109-118.

Llull, J. (2016). Understanding international migration: evidence from a new dataset of bilateral stocks (1960–2000). *SERIEs*, *7*(2), 221–255.

Lucas, R.E.B. (2019). Migration and Development The Role for Development Aid, https://eba.se/en/rapporter/migration-and-development-the-role-for-development-aid-research-overview/11211/

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2018). Political Regime Characteristics and Transitions, 1800-2017. *Center for Systemic Peace*. Retrieved from www.systemicpeace.org

Marshall, M., Gurr, T. R., & Harff, B. (2018). PITF - State Failure Problem Set: Internal Wars and Failures of Governance, 1955-2017.

Martin, Phillip; Taylor, E. (1996). The anatomy of a migration hump. In *Development Strategy, Employment, and Migration: Insights from Models*. (pp. 43–62). Paris: Organization for Economic Cooperation and Development: Organisation for Economic Co-operation and Development ; OECD Publications and Information Center [distributor].

Massey, D. S. (1988). Economic Development and International Migration in Comparative Perspective. *Population and Development Review*, *14*(3), 383–413. https://doi.org/10.2307/1972195

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

Mayer, T., & Zignago, S. (2005). *Market access in global and regional trade*.

McFadden, D., & Zarembka, P. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105–142.

McKenzie, D., & Rapoport, H. (2010). Self-selection patterns in Mexico-US migration: the role of migration networks. *The Review of Economics and Statistics*, *92*(4), 811–821.

Mendola, M. (2012). Rural out-migration and economic development at origin: A review of the evidence. *Journal of International Development*, *24*(1), 102–122.

Ortega, F., & Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies*, *1*(1), 47–74.

Özden, Ç., Parsons, C. R., Schiff, M., & Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960–2000. *The World Bank Economic Review*, *25*(1), 12–56.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, *74*(4), 967–1012.

Rojas-Romagosa, H., & Bollen, J. (2018). Estimating migration changes from the EU's free movement of people principle. *CPB Netherlands Bureau for Economic Policy Analysis*, *385*.

Samuelson, P. A. (1948). International trade and the equalisation of factor prices. *The Economic Journal*, *58*(230), 163–184.

Santo Tomas, P., Summers, L., & Clemens, M. (2009). *Migrants Count: Five Setps Toward Better Migration Data*. Washington, DC: Center for Global Development.

Sasabuchi, S. (1980). A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika*, *67*(2), 429-439.

Sen, A. (2001). What is development about. *Frontiers of Development Economics*, 506–513.

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

Silva, J. M. C. S., & Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, *112*(2), 220–222.

Stark, O. (2006). Inequality and migration: A behavioral link. *Economics Letters*, *91*(1), 146–152.

Tinbergen, J. J. (1962). Shaping the world economy; suggestions for an international economic policy.

UN DESA. (2017). Trends in International Migrant Stock: The 2017 Revision (United Nations database, POP/DB/MIG/Stock/Rev.2017).

United Nations Department of Economic and Social Affairs Population Division. (2006). Trends in Total Migrant Stock, 1960-2000, 2005 Revision.

United Nations Department of Economic and Social Affairs Population Division. (2008). United Nations Global Migration Database. Retrieved from http://esa.un.org/unmigration

Vanderkamp, J. (1971). Migration flows, their determinants and the effects of return migration. *Journal of Political Economy*, *79*(5), 1012–1031.

Zelinsky, W. (1971). The hypothesis of the mobility transition. *Geographical Review*, 219–249.

Appendix

Table A. 1   Overview of the 180 origin and destination countries in the panel data set

ISO code	Country name	ISO code	Country name
ABW	Aruba	DEU	Germany
		DJI	Djibouti
AGO	Angola	DMA	Dominica
ALB	Albania	DNK	Denmark
ARE	United Arab Emirates	DOM	Dominican Republic
ARG	Argentina	DZA	Algeria
ARM	Armenia	ECU	Ecuador
ATG	Antigua and Barbuda	EGY	Egypt, Arab Rep.
AUS	Australia	ESP	Spain
AUT	Austria	EST	Estonia
AZE	Azerbaijan	ETH	Ethiopia
BDI	Burundi	FIN	Finland
BEL	Belgium	FJI	Fiji
BEN	Benin	FRA	France
BFA	Burkina Faso	GAB	Gabon
BGD	Bangladesh	GBR	United Kingdom
BGR	Bulgaria	GEO	Georgia
BHR	Bahrain	GHA	Ghana
BHS	Bahamas	GIN	Guinea
BIH	Bosnia Herzegovina	GMB	Gambia, The
BLR	Belarus	GNB	Guinea-Bissau
BLZ	Belize	GNQ	Equatorial Guinea
BMU	Bermuda	GRC	Greece
BOL	Bolivia (Plurinational State of)	GRD	Grenada
BRA	Brazil	GTM	Guatemala
BRB	Barbados	HKG	Hong Kong SAR, China
BRN	Brunei Darussalam	HND	Honduras
BTN	Bhutan	HRV	Croatia
BWA	Botswana	HTI	Haiti
CAF	Central African Republic	HUN	Hungary
CAN	Canada	IDN	Indonesia
CHE	Switzerland	IND	India
CHL	Chile	IRL	Ireland
CHN	China	IRN	Iran (Islamic Republic of)
CIV	Côte d'Ivoire	IRQ	Iraq
CMR	Cameroon	ISL	Iceland
COD	Democratic Republic of Congo	ISR	Israel
COG	Republic of Congo	ITA	Italy
COL	Colombia	JAM	Jamaica
COM	Comoros	JOR	Jordan
CPV	Cabo Verde	JPN	Japan
CRI	Costa Rica	KAZ	Kazakhstan
CUW	Curacao	KEN	Kenya
CYM	Cayman Islands	KGZ	Kyrgyzstan
CYP	Cyprus	KHM	Cambodia

ISO code	Country name	ISO code	Country name
CZE	Czech Republic	KNA	Saint Kitts and Nevis
KWT	Kuwait	KOR	Republic of Korea
LAO	Lao PDR	LBN	Lebanon
LBR	Liberia	LCA	St. Lucia
LKA	Sri Lanka	SEN	Senegal
LSO	Lesotho	SGP	Singapore
LTU	Lithuania	SLE	Sierra Leone
LUX	Luxembourg	SLV	El Salvador
LVA	Latvia	SRB	Serbia
MAC	Macao SAR, China	STP	São Tomé and Príncipe
MAR	Morocco	RWA	Rwanda
MDA	Republic of Moldova	SAU	Saudi Arabia
MDG	Madagascar	SDN	Sudan
MDV	Maldives	RUS	Russian Federation
MEX	Mexico	SUR	Suriname
MKD	North Macedonia	SVK	Slovak Republic
MLI	Mali	SVN	Slovenia
MLT	Malta	SWE	Sweden
MMR	Myanmar	SWZ	Eswatini
MNE	Montenegro	SXM	Sint Maarten, Dutch part
MNG	Mongolia	SYC	Seychelles
MOZ	Mozambique	SYR	Syrian Arab Republic
MRT	Mauritania	TCA	Turks and Caicos Islands
MUS	Mauritius	TCD	Chad
MWI	Malawi	TGO	Togo
MYS	Malaysia	THA	Thailand
NAM	Namibia	TJK	Tajikistan
NER	Niger	TKM	Turkmenistan
NGA	Nigeria	TTO	Trinidad and Tobago
NIC	Nicaragua	TUN	Tunisia
NLD	Netherlands	TUR	Turkey
NOR	Norway	TZA	United Republic of Tanzania
NPL	Nepal	UGA	Uganda
NZL	New Zealand	UKR	Ukraine
OMN	Oman	URY	Uruguay
PAK	Pakistan	USA	United States of America
PAN	Panama	UZB	Uzbekistan
PER	Peru	VCT	Saint Vincent and the Grenadines
PHL	Philippines	VEN	Venezuela (Bolivarian Republic of)
POL	Poland	VGB	British Virgin Islands
PRT	Portugal	VNM	Vietnam
PRY	Paraguay	YEM	Yemen, Rep.
PSE	West Bank and Gaza	ZAF	South Africa
QAT	Qatar	ZMB	Zambia
ROU	Romania	ZWE	Zimbabwe

Table A. 2   Overview of the main variables used in the analyses, its definitions and sources

Variable	Definition	Source
GDP per capita	The ratio of Purchasing Power Parity (PPP)-adjusted total Gross Domestic Product (GDP) in constant 2011 US dollars, to the total population count.	Penn World Tables, version 9.1.
Age dependency ratio	The ratio of the number of people younger than 15 or older than 64 (dependents) to the working-age population (ages 15-64).	World Development Indicators, World Bank.
Population density	Midyear population divided by land area in square kilometers.	World Development Indicators, World Bank.
Population (total)	The mid-year estimate of all residents, regardless of legal status or citizenship.	World Development Indicators, World Bank.
Polity IV index	This index (Marshall, Gurr, & Jaggers, 2018) considers a nation as strongly democratic if citizens have the ability to express their preferences about policies and leaders through institutions and procedures, executive power is institutionally constrained, and civil liberties are guaranteed. 'Strong' autocracies, on the other hand, are characterized by the presence of sharp restrictions on, or suppression of, competitive political participation. This index ranges from -10 (strongly autocratic) to +10 (strongly democratic).	Center for Systemic Peace (CSP)
Conflict duration	The number of months any origin country has been in any sort of conflict. The types of conflict considered include wars between governments and minorities (ethnic wars) or political challengers (revolutionary wars), and events involving the implementation of policies resulting in the deaths of a significant portion of communal or politicized groups in the total population (genocides and politicides; cf. M. Marshall, Gurr, and Harff, 2018).	CSP's Political Instability Task Force (PITF) data set; authors' estimates.
Natural disaster occurrences	The number of biological, climatological, geophysical, hydrological and meteorological disasters having occurred in a given decade.	EM-DAT database, Université Catholique de Louvain's Centre for Research on the Epidemiology of Disasters (cf. Guha-Sapir, 2019).

Table A. 3   Correlation matrix of selected origin-time variables, including log GDP per capita

	GDPpc at origin	Population at origin	Age dependency ratio at origin	Polity IV index	Number of natural disasters	Networks	Conflict duration	Population density
GDP per capita	1.000							
Population	-0.023	1.000						
Age dependency ratio	-0.780	-0.115	1.000					
Polity IV index	0.409	0.113	-0.435	1.000				
Number of natural disasters	0.024	0.618	-0.147	0.166	1.000			
Networks	0.015	-0.029	-0.016	0.011	-0.018	1.000		
Conflict duration	-0.220	0.310	0.146	-0.109	0.253	-0.013	1.000	
Population density	0.150	-0.011	-0.203	-0.017	0.001	0.000	-0.004	1.000

Table A. 4   Summary statistics of explanatory variables

	N	Mean	St. Dev	Min	Max	Skewness
Migration rate	169271	.001	.021	0	7.801	311.084
Migration flow	169271	1492.9	28612.7	0	4705677	84.469
Ln GDP pc	169271	8.574	1.198	6.126	12.541	.219
GDP pc	169271	11046.39	19331.3	457.506	279000	7.253
Polity IV index	138578	.529	7.149	-10	10	.065
Pop density	156898	266.144	1356.936	.823	21389.1	11.062
Age dep ratio	163524	74.797	19.889	16.856	120.41	-.091
War duration	169271	10.506	29.657	0	120	2.823
Nat. disast. occ.	169271	12.092	26.373	0	284	5.795

Table A. 5   Output Sasabuchi test

	Lower bound	Upper bound
Interval	5.973	12.541
Slope	0.967	-2.404
t-value	3.252	-6.065
P > |t|	0.001	0.000

The overall test of the presence of an inverted U-shape yields a t-value of 3.25, where P > \|t\| = .00574. The extreme point lies at $\ln(GDPpc_{it})$ = 7.85788.

Table A. 6 Base model estimated on a migration data set excluding small island states

	Migration flow
Ln GDPpc orig. ($t-1$)	4.173***
	(0.872)
Ln GDPpc orig. sq. ($t-1$)	-0.267***
	(0.0497)
Ln pop. orig.	-0.133
	(0.279)
Destination-time FE	Yes
Country-Pair FE	Yes
Origin FE	Yes
Year sample	1960-2020
Observations	79,577
Pseudo R-squared	0.901

Robust clustered standard errors in parentheses. ** p<0.01, ** p<0.05, * p<0.1

Note: Small island states are defined as islands with a population of less than 3mln.

Table A. 7 Results from the estimation of the base model with a cubic term

	Migration flow
Ln GDPpc orig. (t – 1)	0.118
	(5.694)
Ln GDPpc orig. sq. (t – 1)	0.209
	(0.659)
Ln GDPpc orig. cubed (t – 1)	-0.0182
	(0.0252)
Total population	-0.0813
	(0.273)
Destination-time FE	Yes
Country-Pair FE	Yes
Origin FE	Yes
Observations	89,490
Pseudo R-squared	0.902

Robust clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
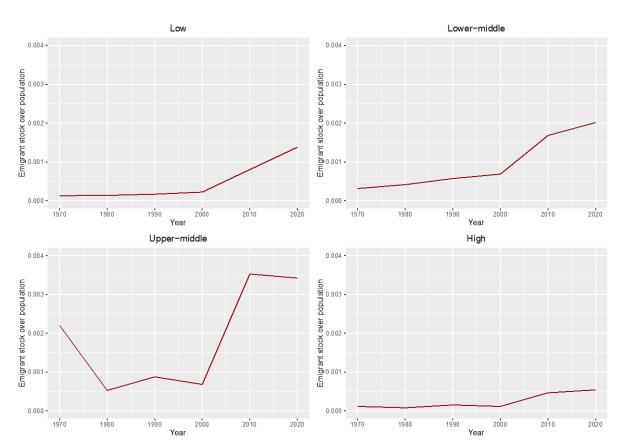
Figure A. 1 Mean bilateral emigration rates over time for countries in each income group (as defined by the World Bank) to countries in all other income groups, in the 1960-2020 timeframe