I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17572

Gravity Predictions of International Migration Flows

Jesús Fernández-Huertas Moraga
Gonzalo López Molina

DECEMBER 2024

DISCUSSION PAPER SERIES

IZA DP No. 17572

Gravity Predictions of International Migration Flows

Jesús Fernández-Huertas Moraga
Universidad Carlos III de Madrid and IZA

Gonzalo López Molina
Analistas Financieros Internacionales

DECEMBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Gravity Predictions of International Migration Flows*

What is the future of international migration flows? The growing availability of bilateral international migration data has resulted in an improved understanding of the determinants of migration flows through the estimation of theory-based gravity models. However, the use of these models as a prediction tool has remained a mostly unexplored research area. This paper estimates simple gravity models of bilateral migration flows for the whole world and projects these models into the future. Our results confirm a limited role for economic factors and a large one for demographic ones, in line with the literature. As a novel contribution, we show that estimates based on net flows are substantially lower than those based on gross flows. The reason is that network effects are historically more correlated with gross than with net flows.

JEL Classification: F22, J11, J61, O15
Keywords: international migration, prediction, gravity model

Corresponding author:
Jesús Fernández-Huertas Moraga
Universidad Carlos III de Madrid
Madrid, 126
28903 Getafe (Madrid)
Spain
E-mail: jesferna@eco.uc3m.es

* The authors are grateful to Simone Bertoli, Jesús Crespo Cuaresma, Clément Imbert and to participants in seminar presentations at Universidad de Alicante, Universidad Pablo de Olavide and AIReF and at the 17th AFD-World Bank Internacional Conference on Migration and Development (Bologna), the Development Economics Workshop (Saint-Etienne), the Workshop on Migration, Climate Change and Economic Development (Universidad de Zaragoza), the 43rd Simposio of the Spanish Economic Association (Madrid), and the VI Regional Modelling Workshop (Sevilla) for their comments and useful discussions, and to Ricard Grebol, Salvatore Di Salvo, Stefano Tarroni and Joan Mut for providing careful research assistance; Jesús Fernández-Huertas Moraga gratefully acknowledges the support of grants CEX2021-001181-M and PID2022-139510NB-I00, nanced by MICIU/AEI/10.13039/501100011033, and Comunidad de Madrid, grant EPUC3M11 (V PRICIT).

1 Introduction

National and international institutions regularly update population projections, needed to predict the future of macroeconomic variables and different policy scenarios. These projections are based on different assumptions on the path of fertility, mortality and, also, immigration and emigration. In fact, international migration flows accounted for 58 per cent of rich countries' population growth between 1990 and 2020.[1] Hence, if the past is any indication, any model of the evolution of the population of a country needs to carefully examine the role of international migration.

The immigration flows into and emigration flows out of a given country depend on the evolution of economic and political conditions both in that particular country and in all countries in the world. The reason is that the rest of the world is a potential destination for the inhabitants of any given country. Despite being aware of this reality, international institutions performing immigration projections tend to use notably rough assumptions. For example, United Nations (2022) collected data mostly from censuses to estimate the number of immigrants living in a particular destination. In the absence of data for some periods and corridors, they generated migration projections as: "... estimates of net international migration were autocorrelated processes centered at zero, restricted such that the values did not exceed plus-or-minus ten per cent of the population size." For the future, United Nations (2022) "... assumed that levels estimated for the period prior to the start of the COVID-19 pandemic (i.e., pre-2020), if stable, would continue through the remainder of the century."

The contribution of this paper is to use the gravity framework to predict the evolution of international migration flows by origin, destination and time. To this end, we both collect data on net flows every 5 years by origin and destination from the stocks generated by United Nations (2020) and data on gross flows offered by Abel and Cohen (2019), and updated in Abel (2022). These data on gross flows are estimates under different assumptions that are consistent with the data on net flows from United Nations (2020). We first regress the migration data on a simple set of controls: demographic variables, networks and GDP per capita as a proxy for economic development. Then, we predict migration flows by using existing projections of these variables: demographic projections by country from United Nations (2022) and GDP per capita from the Shared Socioeconomic Pathways (SSPs) generated by the Intergovernmental Panel on Climate Change and quantified by Riahi et al. (2017),

[1]Own calculations on data from United Nations (2022).

Rogelj et al. (2018) and Gidden et al. (2019).

Contrary to the claims in Beyer et al. (2022), the predictive power of our gravity model with dyadic fixed effects is comparable to that of Bayesian models, such as Welch and Raftery (2022), or Kluge et al. (2024) for the period 2015-2020. In addition, our bootstrapped standard errors are smaller than those associated to Bayesian forecasts. Dyadic fixed effects absorb time-invariant variables such as physical distance, linguistic distance, cultural distance, contiguity, common colonial history, etc. but they also encompass average effects of other variables, including bilateral migration policies. Still, there is enough variation in the data to identify a meaningful role for demography, networks and economic activity.

Our main results reveal a limited quantitative role for economic factors, though, and a large one for demographic ones, in line with the literature, for the evolution of future migration flows. Almost half of the growth in international migration flows that can be expected until 2050 can be attributed to population growth at origin according to our baseline model. What is new in our paper is that we show that predictions based on net flows are substantially lower than those based on gross flows. The reason for this is that we find that network effects, defined as the influence of the stock of migrants from an origin to a destination on future migration flows, are historically more correlated with gross than with net flows. This makes sense because large migration stocks from a particular origin into a particular destination are likely to both attract new immigrants and generate larger return migration flows as well. These two effects are conflated if we use net flows to approximate gross flows.

We are of course not the first ones that have tried to predict immigration and emigration flows in the academic literature, setting aside the examples in international institutions mentioned above.[2] Hanson and McIntosh (2016) can be credited with starting this literature in economics. However, there are two key differences in their approach. First, they estimate their model only on 2000-2010 data on net migration to the OECD. Second, their estimation assumptions are much more restrictive than ours. Specifically, we take the critique by Bertoli (2017) seriously and do not assume that the migration rate of a given cohort-gender cell only depends on the size of these cells at origin and at destination. The rationale for this was that there is economic competition within cells but not across cells, as in Borjas (2003).

[2] A very early version of this paper was specialized to the Spanish case by Fernández-Huertas Moraga and López Molina (2018).

Also, differently from Hanson and McIntosh (2016), we do not assume that origin variables have the same elasticity as destination variables. We will estimate distinct coefficients for origin and destination variables. For example, this gives us the possibility of considering non-linear origin income effects on emigration (Clemens, 2014). Campos (2017) follows the same approach as Hanson and McIntosh (2016) but he extends the dataset to a longer period, using the data from Özden et al. (2011) at the cost of not being able to differentiate by cohorts and gender. A more comprehensive effort than the one undertaken here is the one by Dao et al. (2021). They create and calibrate a general equilibrium model to forecast immigration and emigration until 2100. In their model they distinguish between two different types of workers (college educated and less educated), which we are not able to do. They can do it because they only use 2010 stocks from DIOC-E to parameterize their model. They use the parameterized model to make projections both into the future and into the past. Contrary to them, we adopt a partial equilibrium approach and neglect feedback effects of immigration on economic conditions. More recent prediction efforts in economics have been devoted to the short term forecasting of asylum flows (Boss et al., 2024), using online searches (Böhme et al., 2020), or even the immigration consequences of a change of government in the US (Beine et al., 2024), using migration intentions. Our focus in this paper is more oriented towards the medium and long term.

Given our use of the SSPs to project economic conditions, we are also related to the large literature in economics that looks into the effects of climate change on migration flows (Backhaus et al., 2015; Beine and Parsons, 2015; Desmet and Rossi-Hansberg, 2015; Cattaneo and Peri, 2016; Bertoli et al., 2021; Burzyński et al., 2021, 2022). We focus in the paper on the effects of climate change through GDP changes since our baseline model does not feature any direct effect of climatic factors on migration flows once we control for country-specific time trends following the specification in Burke et al. (2015).

The paper proceeds as follows. Section 2 explains the basic random utility maximization model and how its aggregation results in a classical gravity equation that we both take to the historical data and then project into the future. We next briefly develop the characteristics of the datasets we employ in the estimation and in the projections in section 3. In section 4, we describe our estimation results. Next, in section 5, we present our predictions. We offer some concluding remarks in the last section of the paper, section 6.

2 The Model

In economics, migration decisions are typically modeled as the result of a choice of destination by utility-maximizing individuals (Beine et al., 2016). Given that the choice of destination is a discrete one, the workhorse model for migration studies is the random utility maximization model, developed by McFadden (1974). Once the individual decisions from the random utility maximization model are aggregated, we end up with a classical gravity equation. Next, we briefly develop how to derive the gravity framework.

We start by letting pop_{ot} represent the stock of the population residing in country o at time t. We can then write the scale m_{odt} of the migration flow from country o to country d at time t as:

$$m_{odt} = p_{odt} pop_{ot-1} \tag{1}$$

The term p_{odt} is the probability that an individual from country o migrates to country d at time t, also known as the emigration rate. The random utility maximization model is used to estimate these emigration rates by obtaining the expected value of p_{odt}.

The utility that one individual i who was located in country o at time $t-1$ derives from opting for country d at time t is:

$$U_{iodt} = \beta' x_{odt} + \epsilon_{iodt} \tag{2}$$

The vector x_{odt} includes all deterministic components of utility while ϵ_{iodt} is an individual-specific stochastic component. The distributional assumptions on ϵ_{iodt} determine the expected probability $E(p_{odt})$ that opting for country d represents the utility-maximizing choice (McFadden, 1974). The use of very general distributional assumptions for ϵ_{iodt} leads to this type of expression:

$$\ln\left(\frac{p_{odt}}{p_{oot}}\right) = \frac{1}{\tau}\beta' x_{odt} - \beta' x_{oot} + MRM_{odt} \tag{3}$$

MRM_{odt} is the multilateral resistance to migration term (Bertoli and Fernández-Huertas Moraga, 2013). It reflects the effect of alternative destinations on bilateral migration rates. The parameter τ is the dissimilarity parameter, which is related to the inverse of the correlation in ϵ_{iodt} across alternative destinations. When the independence of irrelevant alternatives holds, then $\tau = 1$ and $MRM_{odt} = 0$. In that case, ϵ_{iodt} is modeled as following an iid extreme value type I distribution and this is what many studies estimate:

$$y_{odt} \equiv \ln\left(\frac{m_{odt}}{m_{oot}}\right) = \beta'(x_{odt} - x_{oot}) + \xi_{odt} \tag{4}$$

After estimating this, we can project out of the classic gravity framework. We would typically take the fitted values $\widehat{y_{odt}}$:

$$\widehat{y_{odt}} = \widehat{\beta}'\left(x_{odt} -' x_{oot}\right) \tag{5}$$

We would transform them into migration flows from origin o to destination d in period t with:

$$\widehat{m_{odt}} = \frac{e^{\widehat{y_{odt}}}}{1 + \sum_{d'} e^{\widehat{y_{od't}}}} \widehat{pop_{ot-1}} \tag{6}$$

This naive approach would generate two types of problems:

- You depend on the correct estimation of $\widehat{pop_{ot-1}}$ to have a good predictor for $\widehat{m_{odt}}$.
- Jensen's inequality. The expectation of a log is not equal to the log of the expectations:

$$E\left[m_{odt}\right] \geq E\left[e^{y_{odt}}\right] E\left[m_{oot}\right] \tag{7}$$

  This can lead to an average underestimation of migration flows if you predict directly out of equation (6).

To solve this issue, known as the retransformation problem by the old health economics literature (Manning, 1998), one option is to separate the emigration decision from the destination choice decision, as Welch and Raftery (2022) do. In this paper, we opt for fitting migration flows directly through a Poisson Pseudo Maximum Likelihood (PPML) estimator, as suggested by Santos Silva and Tenreyro (2006).

Our gravity specification is:

$$m_{odt} = p_{odt} pop_{ot-1} = \exp\left(\beta_d' x_{odt} + \beta_o' x_{oot}\right) \eta_{odt} \tag{8}$$

The vector x_{oot} includes $\log pop_{ot-1}$, which could also affect the emigration rate and hence its coefficient should not be constrained to 1.

The choice of the rest of independent variables included in the model is grounded in the literature. Firstly, many authors have emphasized the role of demography in shaping international migration flows.[3] Young countries have larger migration propensities because migration is an investment decision (Sjaastad, 1962) whose rewards are typically easier to reap when the migratory move takes place at younger ages. As far as destination countries are concerned, the demographic structure may imply that the labor market competition is stronger if there are larger young cohorts at destination and less fierce otherwise (Hanson and McIntosh, 2016). Secondly, economic conditions both at origin and at destination are included in the model, allowing for different elasticities, as a microfounded gravity model will typically imply that the elasticity of migration flows with respect to economic conditions at destination should be larger than with respect to economic conditions at origin. The reason for this is that destinations are more interchangeable among themselves than origins from the point of view of the potential migrant (Beine et al., 2016). Finally, a third element that is added is the role of networks, also called diasporas (Beine et al., 2011). The stocks of immigrants from the same origin present at a destination facilitate migration movements both due to the possibility of reducing migration costs (Mckenzie and Rapoport, 2007) and the possibility of increasing the earnings potential of immigrants at destination (Munshi, 2003).

Many of the main determinants of international migration flows are not easy to project into the future (Beine et al., 2016). The main example would be policies, which have been shown to be quantitatively very relevant in explaining international migration (Bertoli and Fernández-Huertas Moraga, 2015). The implicit assumption of this model with respect to policy is that its average effect during the estimation period at the bilateral level would continue into the future. However, for example, the visa policy of the European Union regarding Turkey might become the most influential factor explaining the arrival of immigrants into Europe within the next 50 years. This is clearly a limitation of this paper and of other existing projections[4] but it is still useful to understand the forces over whose future evolution we do have some information, such as demography. At least in the medium run, we can be reasonably certain about the demographic pressures that different countries will experience.

Beyond the omission of potentially relevant variables, a fundamental assumption of the

[3] Hanson and McIntosh (2010) and Hanson and McIntosh (2012) are two early examples.

[4] Docquier and Machado (2017) would be an exception to this. They do assess the sensitivity of their projections to general changes in migration policies.

model is the null effect of immigration and emigration on the standard of living both in receiving and in emigrant-sending countries. Despite the enduring controversies about the labor market effects of emigration,[5] a common characteristic of most of the studies is that these effects, be they slightly positive or slightly negative, tend to cluster around zero,[6] so that this is not as outlandish an assumption as it may appear at first sight.[7] Fundamentally for our purposes, this assumption allows us to treat economic fundamentals as an exogenous variable in the model. It would even allow us to close the model if we did not need the future demographic predictions to generate our forecasts of immigration and emigration flows. We use these forecasts to update the population figures from United Nations (2022), replacing their estimates of net international migration with our own ones of immigration and emigration flows..

In addition to this, multilateral resistance to migration would still bias the coefficients β_d and β_o. They must be interpreted as the summary effect of the direct variables and their correlation with variables in alternative destinations. However, the prediction of m_{odt} would only be affected if the structure of substitutability of alternative destinations varies over time.

Summing up, the vector x_{odt} contains dyadic fixed effects and:

- Demographic structure. We can include the size of different cohorts.
- Economic conditions. We proxy them by the log of the GDP per capita.
- Networks. This is the stock of co-nationals from country o already residing in the country of destination d. This is the only time-varying dyadic variable in the model.

If our interest laid on the effect of each of these variables on migration flows, we would need to concern ourselves with their potential endogeneity. However, since our interest is to predict, this is not such a pressing issue. Still, Hanson and McIntosh (2016) argue that it is reasonable to assume that the demographic structures of the countries are exogenous to current immigration flows at least in a generation's horizon. Also, we will assume that

[5] Borjas (2003) and Ottaviano and Peri (2012) could be considered the more classic references.

[6] Docquier et al. (2014) is a good example in this respect. Most effects of immigration and emigration on wages are between -1 and 1 per cent over a decade.

[7] Contrary to this idea, Benveniste et al. (2021) quantify the implied effects of immigration embedded in the SSPs.

the effect of immigration and emigration on economic activity is close enough to zero as to consider it negligible, justified by the fact that many estimates of the effects of immigration and emigration actually cluster around zero. We would only need to worry about the endogeneity of the network variable, which has often been instrumented by past settlements or by proxies for these past settlements (Beine et al., 2011).

3 Data

3.1 Dependent Variable

The main source for the immigration flows in this paper are the series of immigration stocks by origin and destination every five years between 1990 and 2020 produced by United Nations (2020). These are immigration stocks mostly based on population censuses. These population censuses can be real censuses, typically decennial, population counts that some countries run every five years, or administrative datasets. In some cases, they are just large surveys. They should be considered as an approximation to actual population figures, particularly for 2020, when they are mostly a projection, given the publication date of United Nations (2020). In between censuses, the data are mostly interpolations of real data.

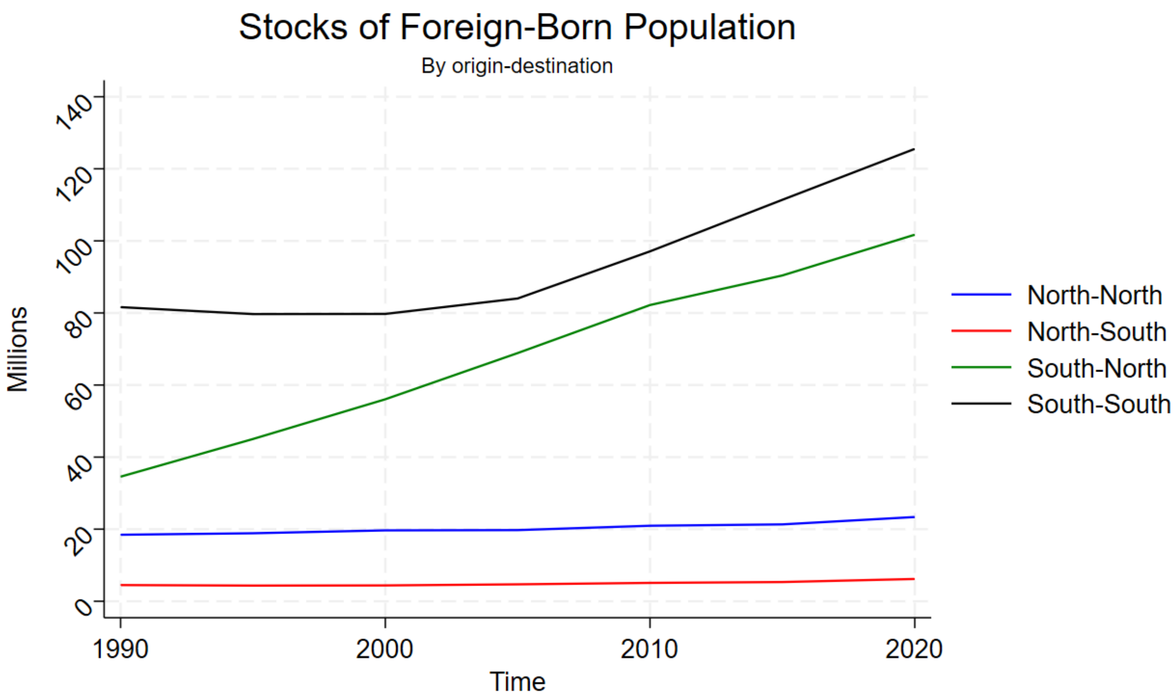
The main advantage of this dataset is how comprehensive it is. It spans the whole world for 30 years with a consistent methodology. This is the largest dataset available in terms of time series and cross-sectionally.[8] The cost is the lack of details about the composition of the flows. We cannot distinguish by age. Furthermore, the random utility maximization model described above implies that m_{odt} should be constructed with data on gross flows. However, the original data will only provide us with net flows that we would have to use as a proxy for gross flows, as in Bertoli and Fernández-Huertas Moraga (2015), for example.

Figure 1 shows the evolution of the migration stocks in the dataset. The figure reflects four series: stocks of emigrants from the South in the North, stocks of emigrants from the South in the South, stocks of immigrants from the North in the South and stocks of immigrants from the North in the North. Following Özden et al. (2011), the North is defined as Western Europe plus the US, Canada, Japan, Australia and New Zealand, while the South is the rest of the world. The actual data will allow us to look into specific bilateral

[8] Özden et al. (2011) created decennial data for 1960-2000 for the whole world. The time period was thus longer, but their data were less frequent and have not been updated.

corridors, but we group countries this way to provide a more general picture of the time series evolution.

Figure 1: Bilateral migration stocks to and from the North and the South (1990-2020)



Source: own elaboration on data from United Nations (2020).

There are two fundamental observations that stand out from looking at figure 1. The first one is how South-South corridors dominate world migration stocks throughout the period. They remained constant in absolute numbers until 2005 below 100 million migrants and then went up until around 125 million migrants by the end of the period. A representative corridor in these bilateral stocks is Indians in the United Arab Emirates, for example, the second largest bilateral corridor in 2020 with 3.5 million individuals. The second observation is the steady increase in emigration from the South to the North in these 30 years, from below 50 million in 1990 to over 100 million migrants in 2020. The effect of the great recession around 2010 on these stocks is barely noticeable as a deceleration in the graph. An example of these stocks would be Mexicans in the US, the largest bilateral corridor in 2020 with almost 11 million individuals. Other than this, North-North corridors are also sizable but stable, while North-South corridors remain very small.

The stocks from figure 1 will be used both to construct net flows and directly as an independent variable to account for the evolution of the network of co-nationals for each origin-destination-time data point.

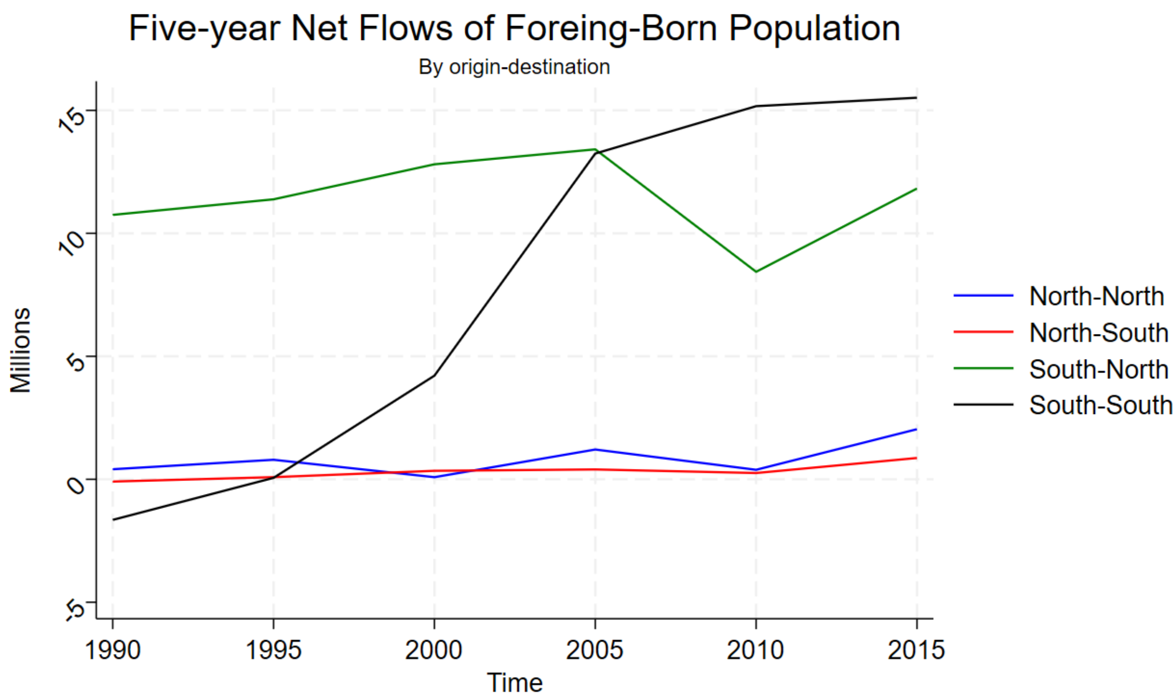
Figure 2 shows the same information as figure 1 but in terms of net flows rather than stocks. For example, the 10 million mark for the South-North migration corridor in 1990 comes from subtracting the 1995 data point from the 1990 one in figure 1. It thus means that immigrants from Southern countries in the North increased by about 10 million between 1990 and 1995. These net flows will be used as a proxy for gross flows when estimating equation (8). In figure 2, we see that South-North net flows actually dominated South-South ones until the 2005-2010 period. After that, coinciding with the great recession, South-South flows grew faster than South-North ones. We can see that South-South flows were actually negative in 1990-1995, but then grew to almost 15 million every 5 years and have remained stable since 2005. South-North net flows also grew, less spectacularly, until 2005-2010, dropped below 10 million in 2010-2015 and then partially recovered above 10 million in 2015-2020. In net flow terms, North-South and North-North corridors confirm in this figure the lack of changes observed in figure 1

The net flows in figure 2 could be consistent with many potential configurations of gross flows. A net flow of zero between countries A and B could be the result of zero immigration but it can also be the result of 10 million migrants of A moving to B while 10 million former A migrants return to B. In theory, the RUM model should be applied to gross flows, as it predicts the probability of moving from a particular origin to a destination at a point in time and net flows are just the aggregation of many of these gross flows. In practice, we tend to only have data on place of birth, which we use as an approximation of the origin o, and place of residence, which we use as an approximation of the destination d. More formally, net migration from country o to country d (nm_{odt}) between periods t and $t+1$ can be obtained as:

$$nm_{odt} = \sum_{r} m^{r}_{odt} - \sum_{r \neq d} m^{d}_{ort} \tag{9}$$

The superscript denotes country of residence at time t while the second subscript would be the residence at time $t+1$. When we use the net flows from figure 2 to approximate gross flows, we are implicitly assuming that $m^{r}_{odt} = 0, \forall r \neq o$, and $\sum_{r \neq d} m^{d}_{ort} = 0$. Furthermore, net flows are typically dropped from the estimation when they are negative so that a lot of

Figure 2: Net migration flows to and from the North and the South (1990-2020)



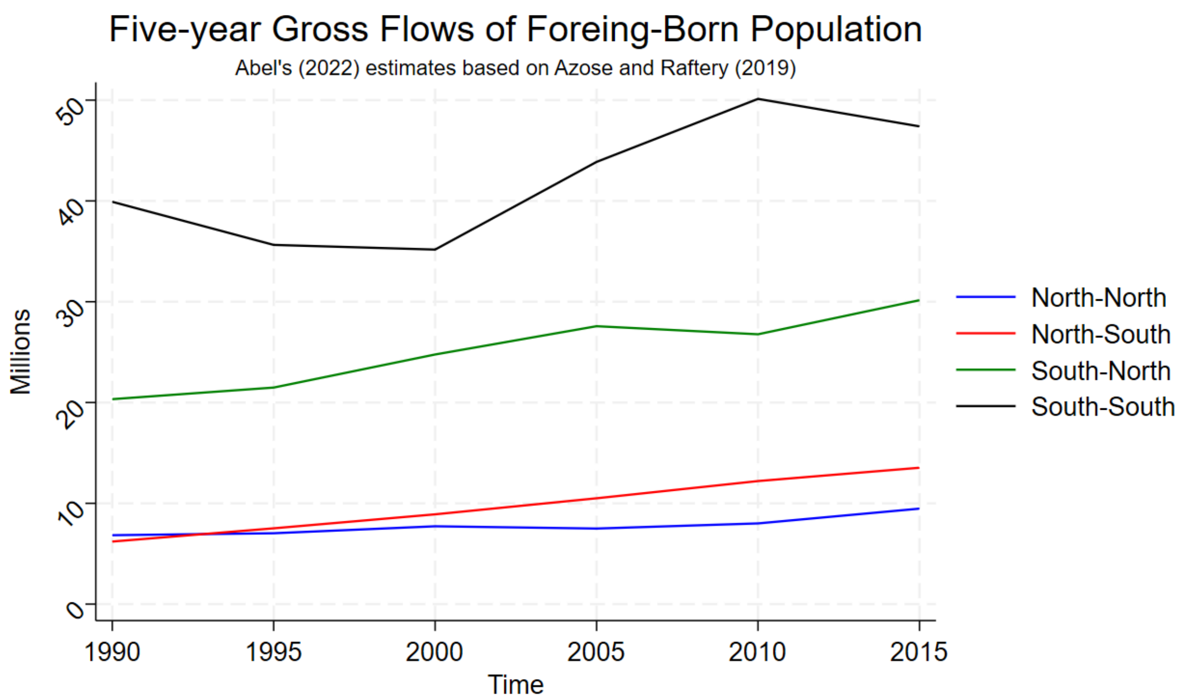
Source: own elaboration on data from United Nations (2020).

information is lost.

In order to correct this, we can use recent estimates of gross international migration flows that have been produced out of the same raw migration data in United Nations (2020). For example, Abel and Sander (2014) estimated gross migration flows by assuming the minimum value of gross flows that would be consistent with computed net flows. In the same line, perhaps the most advanced estimation of gross flows is the one proposed by Azose and Raftery (2019). They generated pseudo-Bayesian estimates of gross flows by taking a weighted mean between the minimum estimates from Abel and Sander (2014) and estimates from a model where independence is assumed between origin-birthplace movements and destination-birthplace movements without origin-destination-birthplace flows. The weight is calculated by fitting the model to 2002-2008 data from the IMEM project on European migration (Raymer et al., 2013).

Figure 3 reproduces figure 2 but with the gross flows based on United Nations (2020) following the methodology in Azose and Raftery (2019) instead of net flows. The data were produced originally by Abel and Cohen (2019) and updated by Abel (2022) with 2020 data. The evolution of international migration flows depicted in figure 3 is more subtle than in figure 2. We can see that South-South flows are dominant with respect to all other corridors and that they grew mostly between 2000 and 2015, with a slight drop in the last five years. For South-North flows, the drop between 2005 and 2015 is now more visible, as well as the recovery in the 2015-2020 period. We can also observe a noticeable increase in gross flows in the South-North corridors from around 20 million migrants in 1990-1995 to almost 30 million in 2015-2020. Finally, while both North-North and North-South migration grew during the period, this growth is more clear in the case of North-South corridors.

Figure 3: Gross migration flows to and from the North and the South (1990-2020)



Source: own elaboration on data from Abel (2022).

3.2 Independent variables

In order to explain the migration data from the previous subsection, we use the historic evolution of population at origin and at destination, and economic conditions at origin and destination. For the population figures, we take total population by origin and destination and its age structure from the World Population Prospects 2022 published by United Nations (2022). For economic conditions, we proxy them by the series on the expenditure side real GDP at chained PPPs to generate GDP per capita from the Penn World Tables (Feenstra et al., 2015): PWT 10.01. We complete the data using PPP GDP per capita series from the World Development Indicators.[9] Missing GDP corridors account for less than 5 per cent of global migration flows.[10] For some of our results, we use some classic time-invariant gravity variables, specifically distance, contiguity, common language and former colonial ties, which we take from Conte et al. (2022).

In some auxiliary regressions, we will also look into the impact of climatic events on migration flows. To this end, following Burke et al. (2015), we take mean temperatures from University of Delaware global gridded monthly land precipitation and temperature (Willmott and Matsuura, 1995; University of Delaware, 2018): version 5.01. For precipitation, also following Burke et al. (2015), we take monthly total precipitations from the Global Precipitation Climatology Centre (Rustemeier et al., 2022).

For the prediction exercise, after estimating the models based on historical data, we need to use projections of our independent variables to be able to forecast future migration flows. For population, we also use the World Population Prospects 2022 (United Nations, 2022). They include 10 population variants: constant fertility, constant mortality, high fertility, instant replacement, instant replacement-zero migration, low fertility, medium, momentum, no change and zero migration. They are depicted in figure 4. For GDP per capita, we take the growth rates implied by the SSPs described by the IGPCC and quantified by Riahi et al. (2017), Rogelj et al. (2018) and Gidden et al. (2019). Our baseline is SSP5 (fossil-fueled development) because we want to follow Burke et al. (2015) once they compute the effects of climate change on GDP, but we need to recall that SSP5 implies reasonably high growth and convergence.

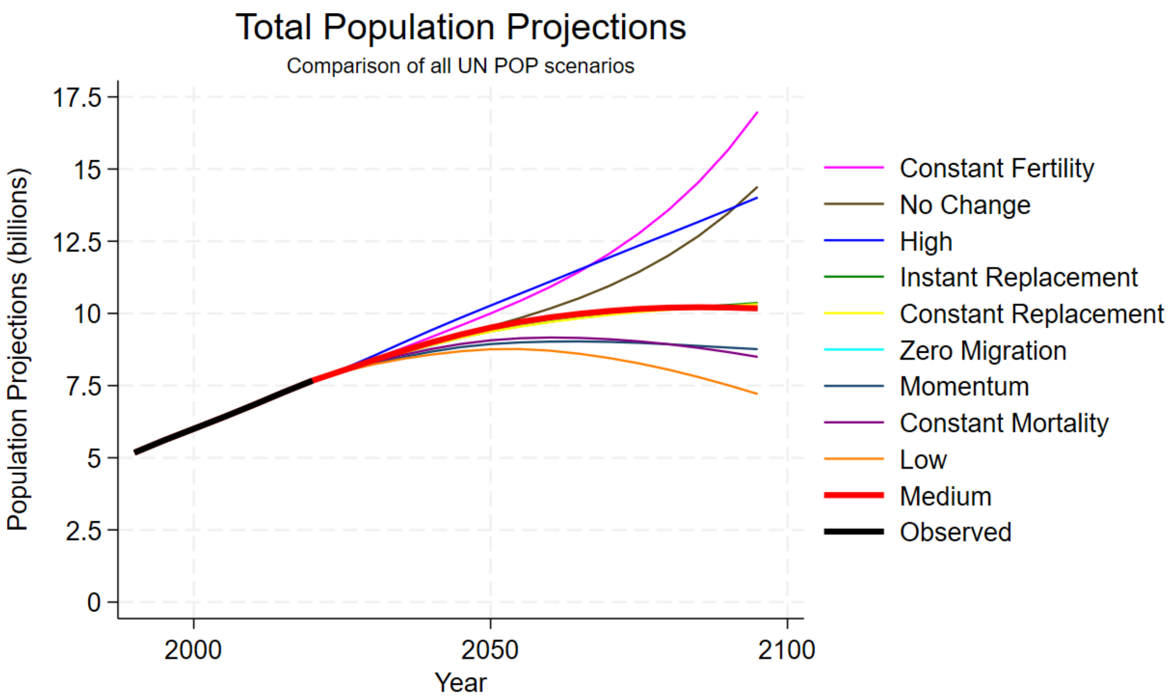
4 Estimation Results

This section summarizes the main estimation results of the paper.

[9]Data from World Bank (2024) accessed on 9-2-2024 using Azevedo (2011).

[10]Appendix A.1 describes how we arrive at our main estimation sample.

Figure 4: World Population according to the WPP 2022



Source: own elaboration on data from United Nations (2022).

4.1 Baseline

We first estimate different versions of equation (8) by sequentially adding explanatory variables in Table 1. We present the results on a homogenous sample where we have data for all of our main explanatory variables. The main source of missing variables, as explained above, is the GDP per capita variable, but results are unaffected for other variables by these missing corridors.[11] Our baseline sample counts 231,324 observations, corresponding to 6 five-year periods that span from 1990 to 2015, 200 countries of origin and 200 countries of destination for the migrants. Summary statistics are presented in Appendix A.1.

In the first column of Table 1, we regress gross migration flows by origin, destination and five-year period on the log of the population at destination and the log of the population at origin plus a full set of origin-destination fixed effects ($200 \times 199 = 39,800$).[12] The elasticity of gross migration flows with respect to the population at destination is -0.58, reflecting a

[11]Results available from the authors upon request.

[12]We use the PPMLHDFE Stata command developed by Correia et al. (2020).

negative association between growing populations at destination and the reception of international migrants, while the elasticity with respect to population at origin is 1.81, significantly larger than 1, which shows that the population at origin, in this simple regression, does not just scale the probability of migration, which would imply a unitary coefficient, but it also affects this probability, with increases in population at origin increasing it as well.

In the second column of Table 1, we add the share of the population aged 15 to 49 as an explanatory variable, both at the origin and at the destination. The objective of adding these variables is to control for the age structure of the population, which we know is relevant for migration decisions, as migrating is an investment decision. The World Population Prospects 2022 (United Nations, 2022) provide us with the full age structure of the population. We choose this particular specification for comparability purposes with the literature. If we add the full population structure, we gain little in terms of the in-sample explanatory power of the model, but the multicollinearity problems among the sizes of different age groups make the models very unstable for prediction purposes. The elasticity of migration flows with respect to the share of young population at destination is positive and significant, while it is negative and marginally significant for the share of young at the origin. This may appear unintuitive at first, but these are elasticities that keep the total population at origin and at destination constant. If we do not control for total population, the share of young people at origin becomes intuitively positive and significant, while the share of young people at destination becomes non-significant.[13]

In the third column of Table 1, we add the logarithm of the migrant stock plus 1, so as not to lose zero stocks, as an independent variable. The elasticity of gross migration flows with respect to this network variable, the stock of migrants from that origin at that destination at the beginning of the period, is positive and significant. The interpretation of the 0.15 coefficient is that an increase of 1 per cent in the number of migrants at a destination increases gross migration flows from that origin to that destination by 0.15 per cent. This is sizable enough to alter significantly the type of predicted migration flows under different models, as we will see in the next section, but much smaller than other estimates in the literature (Beine et al., 2011), particularly those that do not control for dyadic fixed effects. Our dyadic fixed effects absorb a large part of the classic network effect uncovered by the literature in such a short panel.

[13]Results available from the authors upon request.

Table 1: Baseline results: gross flows

	(1)	(2)	(3)	(4)	(5)
Log Population at Destination	-0.577***	-0.887***	-1.178***	-1.118***	-1.119***
	(0.184)	(0.228)	(0.201)	(0.269)	(0.269)
Log Population at Origin	1.808***	2.145***	2.097***	1.918***	1.919***
	(0.257)	(0.285)	(0.285)	(0.266)	(0.263)
Share 15-49 at Destination		2.651***	3.481***	3.085***	3.093***
		(0.904)	(0.846)	(0.962)	(0.935)
Share 15-49 at Origin		-1.960*	-1.872	-1.420	-1.403
		(1.147)	(1.216)	(1.147)	(1.157)
Log (Migrant Stock +1)			0.152***	0.143***	0.143***
			(0.048)	(0.051)	(0.052)
Log GDPpcPPP at Destination				0.196**	0.195*
				(0.096)	(0.105)
Log GDPpcPPP at Origin				-0.093	-0.128
				(0.152)	(0.622)
(Log GDPpcPPP at Origin)^2					0.002
					(0.035)
Observations	231324	231324	231324	231324	231324
Dyadic FE	Yes	Yes	Yes	Yes	Yes
GDP origin p-value					0.828
Pseudo R^2	0.954	0.954	0.955	0.955	0.955

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors by origin country in parentheses. The dependent variable are pseudo-Bayesian estimates of gross flows following Azose and Raftery (2019), provided by Abel (2022). Dyadic FE refers to origin-destination fixed effects. GDP origin p-value from a test that both GDP per capita at origin coefficients are zero.

Our preferred and baseline specification is presented in column 4 of Table 1, where we add the log of the GDP per capita both at origin and at destination. The earlier variables keep their size and significance, while only GDP per capita at destination enters significantly at the 5 per cent level in this specification, with an elasticity of 0.20. The coefficient on the

GDP per capita at origin, negative at -0.09, does not imply a significant correlation with gross migration flows.

Finally, in column 5 of Table 1, we explore whether the lack of significance of the coefficient on the GDP per capita at origin might be coming from the non-linear relationship between development and emigration uncovered by Clemens (2014), and studied by Dao et al. (2018) or Clemens and Mendola (2024) among others. To this end, we add the square of the log of the GDP per capita at origin as an explanatory variable that could pick up the positive correlation between emigration and development for low levels of income at origin. In column 5, the coefficient on the squared term is positive and insignificant, and a joint test of the significance of both GDP at origin coefficients is clearly unable to reject the null that they are both zero.[14]

4.2 Other Specifications

Our baseline results in Table 1 present a very particular choice of fixed effects and misses some classic explanatory variables in a gravity model, notably distance. In Table 2, we check the stability of the coefficients in our model to alternative fixed effects specifications and to the addition of classic time-invariant gravity variables.

First, column 1 in Table 2 reproduces our baseline model, from column 4 in Table 1, in order to make comparisons easier. Next, in column 2 we substitute the dyadic origin-destination fixed effects for simple origin fixed effects and destination fixed effects. Hence, the specification in column 2 of Table 2 features 400 fixed effects' variables, compared to 39,800 fixed effects in our baseline model in column 1. Instead of the dyadic fixed effects, we include four classic variables from the CEPII gravity dataset (Conte et al., 2022): distance between the origin and the destination, contiguity, a dummy for the existence of a common language and a dummy for a former colonial relationship. The gravity variables enter with the expected signs. Migration flows are negatively related to distance and positively related to contiguity, common language and common colonial past. The likelihood of the model, as measured by the Pseudo R^2 , goes down from 0.96 to 0.86, but the rest of variables keep their size, with the notable exception of the network variable, which goes up from 0.14 in our baseline to 0.48 in a model without dyadic fixed effects. This means that a big part of

[14]This does not mean that the relationship is not present in our dataset. It comes back when we use net flows instead of gross flows, as in the cited articles. Results available from the authors upon request.

Table 2: Other specifications: gross flows

	(1)	(2)	(3)	(4)
Log Population at Destination	-1.118***	-1.393***		
	(0.269)	(0.336)		
Log Population at Origin	1.918***	1.741***		
	(0.266)	(0.282)		
Share 15-49 at Destination	3.085***	4.178***		
	(0.962)	(1.202)		
Share 15-49 at Origin	-1.420	-1.461		
	(1.147)	(1.363)		
Log (Migrant Stock +1)	0.143***	0.476***	0.494***	0.199***
	(0.051)	(0.024)	(0.025)	(0.023)
Log GDPpcPPP at Destination	0.196**	0.196*		
	(0.096)	(0.102)		
Log GDPpcPPP at Origin	-0.093	-0.147		
	(0.152)	(0.166)		
Log Distance		-0.261***	-0.271***	
		(0.058)	(0.061)	
Contiguous		0.324***	0.291***	
		(0.099)	(0.091)	
Common Language		0.147*	0.146*	
		(0.076)	(0.077)	
Former Colony		0.572***	0.572***	
		(0.084)	(0.089)	
Observations	231324	219376	219376	231324
Dyadic FE	Yes			Yes
Orig FE, Dest FE		Yes		
Orig-Year FE, Dest-Year FE			Yes	Yes
Pseudo R^2	0.955	0.864	0.900	0.987

* $p<0.1$, ** $p<0.05$, *** $p<0.01$. Clustered standard errors by origin country in parentheses. The dependent variable are pseudo-Bayesian estimates of gross flows following Azose and Raftery (2019), provided by Abel (2022). Dyadic FE refers to origin-destination fixed effects.

a potential network effect is picked up in our baseline model by those dyadic fixed effects.

In column 3 of Table 2, we opt for a more aggressive fixed effects specification, where we control for interacted origin-time and origin-destination fixed effects. Hence, we lose all of our origin-time and destination-time variables that proxy for demographic and economic conditions at origin and at destination, since they are collinear with this fixed effects specification. The coefficients move very little with respect to column 2, which suggests that the bias induced by proxying for origin and destination factors using our three demographic and economic variables is not very large. The Pseudo R^2 only increases from 0.86 to 0.90 with this specification that controls for absolutely everything that can be going on at the origin and at the destination without dyadic components.

Finally, column 4 of Table 2 saturates the model with dyadic, origin-time and destination-time fixed effects. This means that we can only identify the coefficient of the network variable, the only dyadic time-varying variable in the model. There are two elements that must be emphasized for this specification. First, the Pseudo R^2 is not that far from the one in our baseline model, from 0.99 to 0.96, which means that our baseline model is very comparable in terms of likelihood with this more complete model. Second, the elasticity of migration flows with respect to the stock of co-nationals from the same origin in the same destination is 0.20, much closer to the 0.14 of our baseline model, and within its confidence interval, than to that in the models without dyadic fixed effects, such as columns 2 and 3.

Column 4 in Table 2 presents the model with the highest likelihood, but, of course, it cannot be used for prediction purposes, since we cannot predict time effects by origin and by destination into the future. The same is true about column 3.

4.3 Other Dependent Variables

Our baseline results are based on the definition of gross migration flows developed by Azose and Raftery (2019). They create gross flows out of migration stocks using reasonable assumptions, but it is nonetheless interesting to look into the impact of the type of proxy we can use for migration flows on the estimates of the gravity model.

This exercise is performed in Table 3, where we compare our baseline model, from column 4 in Table 1 to the same model run on different versions of the dependent variables. In column 1, we reproduce our baseline model to ease the comparisons. In column 2, our dependent variable is the classic difference in stocks, where negative values are simply turned into zeroes. There are two main differences with our baseline model. First, the log of the migrant stock, our proxy for migration networks, becomes insignificant in the specification for net flows, possibly because return migration, which enters negatively the definition of net flows, is positively related to migration stocks: more people will return when there are larger stocks in a destination. Second, the elasticity of net flows with respect to GDP at destination is also larger. Again, this could be related to the return migration flow, as increases in GDP at destination should be reducing return migration flows.

Columns 3 to 5 in Table 3 present the results of our benchmark specification for three other definitions of the dependent variable. In column 3, negative net flows are added as positive flows in the opposite direction, following Beine and Parsons (2015), but these can also be considered net flows. In columns 4 and 5, we introduce two different assumptions about the construction of gross migration flows out of stock migration data. Abel and Sander (2014) built the minimum gross migration flow that is compatible with observed migration stocks. This is presented in column 4. In column 5, the assumption is the same but the model is closed to account for missing corridors. All of these models turn out to be quite comparable with our baseline model, even if the network variable is not generally significant. Still, it is not significantly different from our baseline estimate of 0.14 either.

For the rest of the paper, we focus on the difference between columns 1 and 2 as the difference between net and gross flows while we comment our results.

4.4 Adding Climatic Variables and Time Trends

When we look at the sensitivity of migration predictions with respect to economic conditions, we will be using the SSPs, as advanced in Section 3. These SSPs were developed to understand the economic effects of different climate change scenarios. We could also think of incorporating directly climatic events in our baseline model in order to directly test the sensitivity of migration flows to climatic variables, as other papers did (Beine and Parsons, 2015; Cattaneo and Peri, 2016).

This is what we do in Table 4. Since we lose observations on small countries when we merge our dataset with the data on temperatures and precipitation (see Appendix A.1), we first show how our baseline model changes in this reduced sample with 167,682 observations in column 1. There are no relevant changes although the coefficient on the GDP at destination is more imprecisely estimated and only significant at the 10 per cent level. The magnitude

Table 3: Other specifications: different dependent variables

	(1) Gross Flows Pseudo-Bayesian	(2) Net Flows Stock Diff Drop Neg	(3) Net Flows Stock Diff Rev Neg	(4) Gross Flows Open Acc	(5) Gross Flows Closed Acc
Log Population	-1.118***	-1.547**	-2.388***	-1.750***	-1.964***
at Destination	(0.269)	(0.723)	(0.535)	(0.622)	(0.613)
Log Population	1.918***	1.430**	3.063***	1.561***	1.648***
at Origin	(0.266)	(0.659)	(0.554)	(0.593)	(0.524)
Share 15-49	3.085***	9.214***	6.484***	8.334***	8.756***
at Destination	(0.962)	(3.343)	(2.057)	(2.622)	(2.748)
Share 15-49	-1.420	-1.868	0.223	-2.791	-2.571
at Origin	(1.147)	(2.076)	(2.480)	(2.810)	(2.369)
Log (Migrant Stock +1)	0.143***	0.004	0.100	0.060*	0.037
	(0.051)	(0.032)	(0.069)	(0.034)	(0.046)
Log GDPpcPPP	0.196**	1.015***	0.257	0.381*	0.672***
at Destination	(0.096)	(0.284)	(0.188)	(0.231)	(0.244)
Log GDPpcPPP	-0.093	-0.033	-0.237	0.109	0.154
at Origin	(0.152)	(0.382)	(0.248)	(0.338)	(0.326)
Observations	231324	231324	231324	231324	231324
Dyadic FE	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.955	0.903	0.888	0.901	0.890

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors by origin country in parentheses. The dependent variables are provided by Abel (2022). Dyadic FE refers to origin-destination fixed effects.

of the rest of coefficients remains virtually unchanged.

In column 2 of Table 4, we look for direct effects of climatic variations, measured by average temperatures and total precipitations per country and time period, on gross migration flows. We find no significant correlation with temperatures at origin, but a strongly significant and positive correlation with temperatures at destination. This indicates that between 1990 and 2020, gross migration flows were more attracted by countries where temperatures

Table 4: Other specifications: adding climatic variables

	(1)	(2)	(3)
Log Population at Destination	-1.138***	-1.261***	-1.167***
	(0.270)	(0.218)	(0.271)
Log Population at Origin	1.919***	1.993***	1.892***
	(0.269)	(0.280)	(0.269)
Share 15-49 at Destination	3.140***	3.332***	2.881***
	(1.007)	(0.837)	(1.000)
Share 15-49 at Origin	-1.679	-1.963	-1.620
	(1.186)	(1.242)	(1.183)
Log (Migrant Stock +1)	0.149***	0.154***	0.150***
	(0.055)	(0.055)	(0.056)
Log GDPpcPPP at Destination	0.192*		0.134
	(0.100)		(0.094)
Log GDPpcPPP at Origin	-0.074		-0.102
	(0.154)		(0.156)
Average Temperature at Destination		0.245***	0.228***
		(0.074)	(0.079))
Average Temperature at Origin		-0.029	-0.030
		(0.099)	(0.099)
Total Precipitation at Destination (x1000)		-0.365	-0.354
		(0.345)	(0.341)
Total Precipitation at Origin (x1000)		-0.263	-0.267
		(0.214)	(0.209)
Observations	167682	167682	167682
Pseudo R^2	0.952	0.952	0.952
Dyadic FE	Yes	Yes	Yes

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors by origin country in parentheses. The dependent variable are pseudo-Bayesian estimates of gross flows following Azose and Raftery (2019), provided by Abel (2022). Dyadic FE refers to origin-destination fixed effects.

were increasing. This correlation is strong enough to drive the coefficient on GDP per capita at destination insignificant in column 3 of Table 4, although the difference is not statistically significant with column 1. Also, the addition of the climatic variables does not change much the likelihood of the models, as the lack of variation in the Pseudo R^2 shows.

One could be concerned about the statistical significance of temperatures at destination in Table 4. Should we add temperatures at destination to our baseline model to predict future migration flows? The significance of this coefficient would argue for the inclusion. However, one could think that climatic variations should also affect economic conditions (Burke et al., 2015), so that a specification with both GDP and temperatures would not have a straightforward interpretations. In order to understand whether temperatures at destination are a robust correlate of migration flows or just the expression of a secular trend, in Table 5 we repeat the same specifications as in Table 4, but with the addition of origin-specific and destination-specific time trends. These models are inspired by the main specification in Burke et al. (2015), who argued that effects of climatic variables, temperatures in particular, beyond the trend could be used to isolate the causal effect of temperature changes on GDP growth.

In Table 5, the addition of origin and destination-specific linear time trends leads to an insignificant estimate for all climatic variables. This is not generally the case for demographic and economic variables. In particular, GDP per capita at destination becomes more significant and even larger with time trends, and the only novelty is that GDP at origin also becomes negatively significant in this specification, signaling that negative deviations from trend in GDP per capita at origin are more correlated with emigration than regular decreases in migration flows. One could be tempted to use our model with linear trend in column 1 of Table 5 as our baseline for predictions in the next section. However, the origin-specific and destination-specific time trends are too imprecisely estimated and too volatile to derive meaningful predictions from them.

Summing up, our joint interpretation of Tables 4 and 5 is that any effect of climatic variables on migration flows should be accommodated in a simple model like ours through their effect on economic conditions. As a result, we will show the sensitivity of the predictions derived from our baseline to changes in GDP associated to different SSPs and we will also compare migration flows in the two main scenarios proposed by Burke et al. (2015): without climate change in SSP5 and with climate change, for which they estimated a decrease in

Table 5: Other specifications: adding climatic variables with country time trends

	(1)	(2)	(3)
Log Population at Destination	-4.591***	-4.597***	-4.462***
	(0.600)	(0.560)	(0.619)
Log Population at Origin	5.769***	6.004***	5.827***
	(1.707)	(1.668)	(1.697)
Share 15-49 at Destination	3.769***	4.184***	3.542**
	(1.424)	(1.458)	(1.418)
Share 15-49 at Origin	-0.434	-0.940	-0.480
	(2.612)	(2.554)	(2.631)
Log (Migrant Stock +1)	0.084	0.102*	0.087
	(0.057)	(0.055)	(0.056)
Log GDPpcPPP at Destination	0.285***		0.267***
	(0.101)		(0.102)
Log GDPpcPPP at Origin	-0.217*		-0.223*
	(0.120)		(0.119)
Average Temperature at Destination		0.076	0.049
		(0.049)	(0.044)
Average Temperature at Origin		0.114	0.112
		(0.081)	(0.080)
Total Precipitation at Destination (x1000)		-0.018	-0.014
		(0.166)	(0.16)
Total Precipitation at Origin (x1000)		-0.126	-0.131
		(0.227)	(0.224)
Observations	167682	167682	167682
Pseudo R^2	0.972	0.972	0.972
Dyadic FE	Yes	Yes	Yes
Orig and Dest linear trends	Yes	Yes	Yes

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors by origin country in parentheses. The dependent variable are pseudo-Bayesian estimates of gross flows following Azose and Raftery (2019), provided by Abel (2022). Dyadic FE refers to origin-destination fixed effects.

world GDP of 23 per cent by 2100 associated to an increase in average global temperatures of 4.3 degrees Celsius.

5 Predictions

In this section, we will take the estimated coefficients from the results presented above and use them to predict the evolution of future migration flows all the way up to 2100. In order to perform this exercise, we take the population structure of origin and destination countries as given, represented in the variable of the share of individuals aged 15 to 49. We also assume that immigration and emigration do not significantly alter GDP projections and take the GDP growth associated to the SSPs also as given. On the contrary, we use our own predictions to update the total population of origin and destination countries and, more relevantly, the migrant stock in each origin-destination pair. We use actual data on migrant stocks for our predictions of migration between 2020 and 2025 and then our updates for the rest of the century. We detail below the assumptions that we use.

From 2025 onwards, we always take the direct prediction from equation (8) for each corridor. We only make sure that the total emigration flow does not exceed the total population of the country of origin. If it does, we just distribute the total population across destinations based on weights drawn from the direct prediction of equation (8). To update migrant stocks, we first assume that the death rate of migrants is the same as the predicted death rate at the origin country according to United Nations (2022). This is not realistic, but it allows us to keep the world population constant in our prediction.[15] Second, we update migrant stocks by adding our predicted net flows by corridor to the surviving migration stock. The hidden assumption here is that the first people to leave a country are the former migrants from the point of view of net migration receivers, while they are natives from the point of view of net migration senders. If gross flows from A to B are larger than gross flows from B to A, the stock of A people in B increases while the stock of B people in A decreases. This is also unrealistic, but it can be considered as a way of generating a lower bound on migration stocks, because they tend to be reduced globally by mortality. Given the relevance of the stock variable for our predictions, as it will be seen, we keep this conservative assumption.

To update the populations of the countries, what we do is to subtract the net migration flows provided by United Nations (2022) for each scenario in the WPP 2022 and add instead our model-generated net migration flows by destination. We then apply the population growth rate provided by United Nations (2022) to the model-generated population of the previous period.

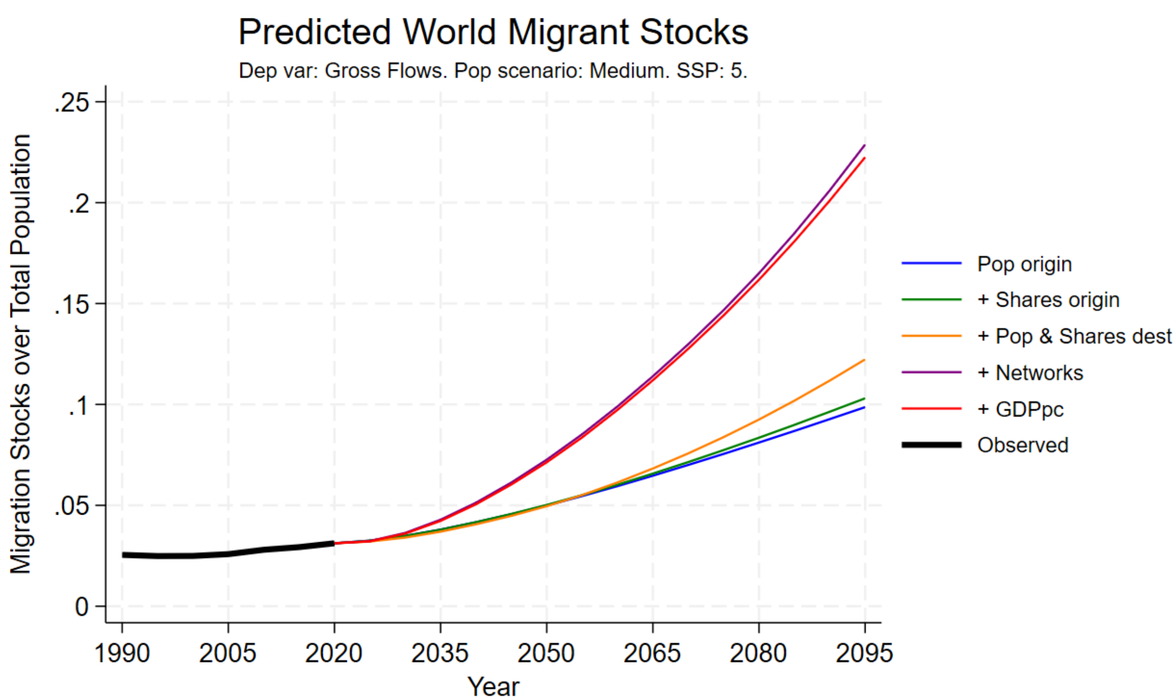
[15] Results are robust to this assumption. Alternative predictions available upon request.

We choose to present our main prediction exercises in terms of the stock of international migrants in the world or in some particular corridor over the total population. This is helpful to fix ideas in a metric that can be widely understood.

5.1 By Model

We start by showing the relative influence of our independent variables for the prediction of future migration flows in Figure 5. The figure is built by predicting migration flows using equation (8) and the procedure describe above to add to existing migration stocks in the world, depicted in solid black for the period 1990-2020. These predicted migration stocks are then divided by the world population according to the Medium population scenario provided by United Nations (2022). GDP per capita is assumed to evolve according to SSP5, a scenario of high growth and convergence across countries. This scenario is known as "Fossil-fueled Development."

Figure 5: Predicted Share of Migrants by Model



Source: own elaboration.

Five different gravity predictions are shown in Figure 5, corresponding to models presented in columns 2, 3 and 4 of Table 1, and two additional models that can be found in the Appendix A.2. The simplest model ("Pop origin" in blue in the figure) includes only dyadic fixed effects and population at origin as the only time-varying regressor. It allows us to see the effect of the size of origin-country populations on predicted migration flows, assuming average migration rates by corridor remain at their 1990-2020 levels, which is the information contained in the fixed effects. Under this simple model, world migration stocks would increase from 3.2 per cent of the world population in 2020 to 5.0 per cent in 2050 and to 9.7 per cent in 2095.

Figure 5 allows us to realize that the population structure at origin does not add much to the prediction, while population at destination only starts to matter after 2060, when predictions should be considered less and less reliable, as they would start applying to populations that have not been born yet.

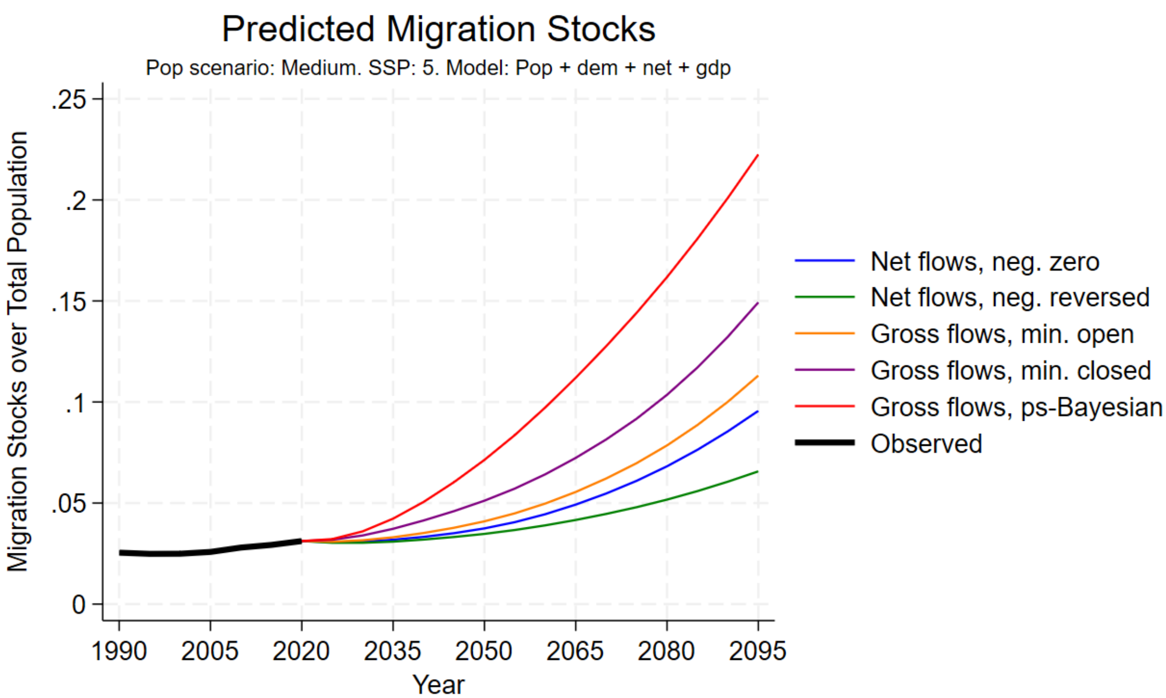
The variable that seems to matter the most for the prediction, beyond population at origin, is the stock of co-nationals residing in the destination country, our proxy for a network component of migration flows (purple line in Figure 5). The inclusion of this variable in the model, even with a relatively low elasticity of 0.15 (column 3 in Table 1) increases the prediction of the stock of migrants in 2050 up to 7.1 per cent of the world population, and up to 22.3 per cent by 2095.

Finally, another element that should be emphasized is the very small impact that including economic variables has on the predictions. Once we include GDP per capita at origin and at destination, the relative convergence of the SSP5 scenario leads to a small reduction of world migration stocks (red line in Figure 5) with respect to a model without economic components. They would reach 7.0 per cent in 2050 and 21.7 per cent by 2095.

5.2 By Dependent Variable

Next, we look into the impact of the choice of dependent variable on the predicted share of migrants. Figure 6 shows the results of this exercise, whose estimates were presented in Table 3. We recall that the main difference between the use of alternative definitions of migration flows was the relevance of the network variable, larger in our baseline model with gross migration flows built according to the pseudo-Bayesian method of Azose and Raftery (2019). This is reflected in the fact that the prediction for our baseline model (red line in Figure 6) is much larger than that of any of the predictions based on the rest of dependent variables. By 2050, our baseline model predicts that 7.0 per cent of the world population will live abroad, up from 3.2 per cent in 2020. If we take instead the simple difference in stocks as our dependent variable, as in column 2 of Table 3, the model predicts that 3.7 per cent of the world population will be international migrants by 2050 (blue line in Figure 6). By 2095, the difference would be between 21.7 per cent with our baseline gross flows and 9.6 per cent with these alternative net flows.

Figure 6: Predicted Share of Migrants by Dependent Variable



Source: own elaboration.

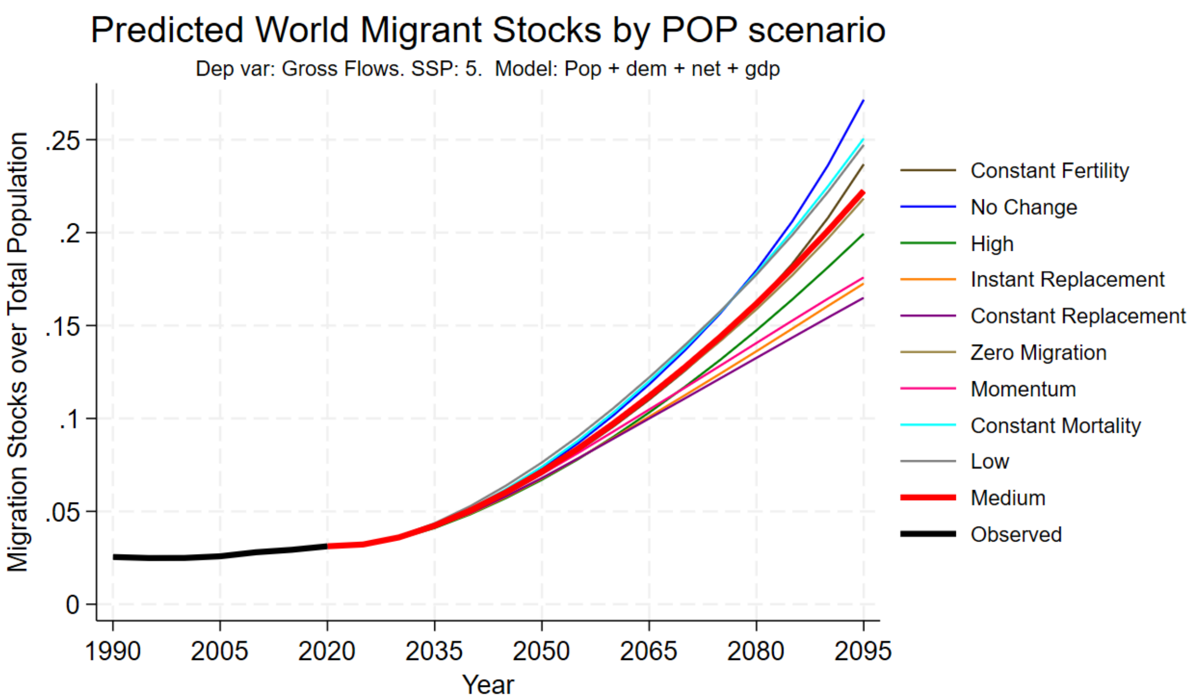
In general, models with gross flows as the dependent variable predict more growth of international migration into the future. The main differences that we see in Figure 6 have to do both with the size of the network coefficient and with that of the elasticity of migration flows with respect to the GDP per capita at destination, which explains, for example why net flows where we turn negative values into zeroes predict more migration that net flows where we reverse negative values (green line). The GDP coefficient is about 4 times larger for the former model (column 2 in Table 3) than in the latter (column 3).

5.3 Sensitivity to Demographic and Economic Scenarios

All of the above predictions were keeping constant the population projections (Medium) and the economic scenario (SSP5). In this subsection, we turn our attention to the sensitivity of the predictions to different demographic and economic scenarios.

Starting with demography, Figure 7 represents how our baseline model changes as we vary the demographic projections that we feed from the future based on the World Population Prospects 2022 (United Nations, 2022). What is more remarkable about Figure 7, which shows our baseline prediction in thick red, is how concentrated all predictions are in the medium term, around 7 per cent as the world migrant share in 2050. In fact, we only see big differences across variants for the end of the century, with all variants predicting world migrant stocks between 15 and more than 25 per cent of the world population.

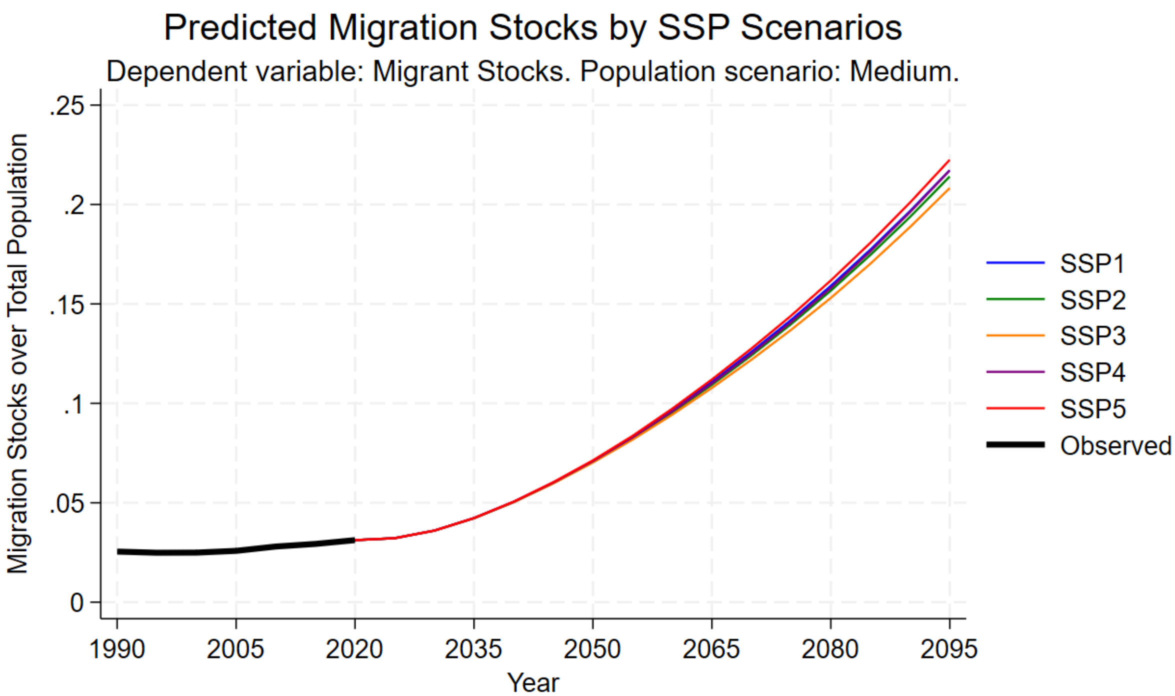
Figure 7: Predicted Share of Migrants by Population Variant



Source: own elaboration.

The differences across socioeconomic scenarios are even less remarkable in our baseline model. Figure 8 represents the predicted migration stocks that result from using the GDPs associated to different SSPs to proxy the evolution of origin and destination countries. Given the low elasticity of migration flows to respect to GDP that was found in column 4 of Table 1, it is not surprising that it is actually hard to distinguish across SSPs in terms of their migration consequences.

Figure 8: Predicted Share of Migrants by SSP



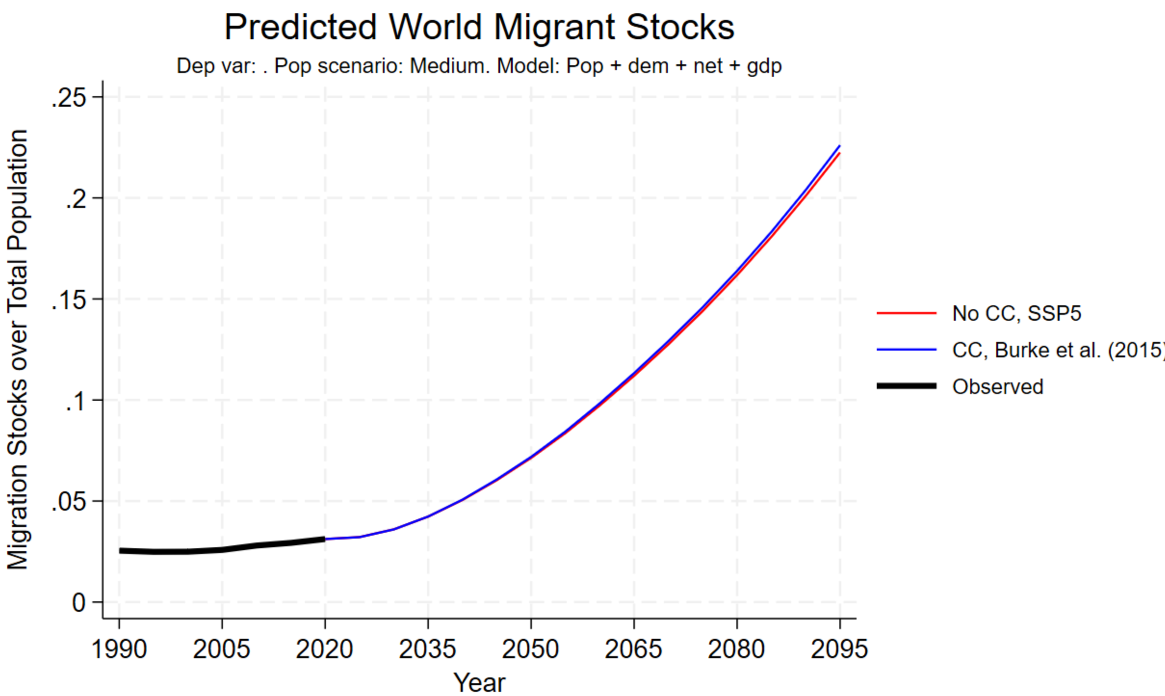
Source: own elaboration.

5.4 Climate Change

In addition to the five SSPs, Figure 9 adds a sixth scenario, based on Burke et al. (2015). This scenario is a modified version of SSP5, where the GDP growth implied by SSP5 is reduced according to the estimates of Burke et al. (2015) for the effects of climate change on economic activity. We can see that a potential reduction of economic activity associated to a warming planet would increase international migration, but not by much if the relationship between GDP changes and migration continues to be the same as it was between 1990 and 2020.

We could be tempted to argue that climate change, here equivalent to large changes in the distribution of GDP growth across countries, has no effect of international migration flows.

Figure 9: Predicted Share of Migrants and Climate Change



Source: own elaboration. CC refers to climate change, defined as the baseline decrease in GDP projected by Burke et al. (2015).

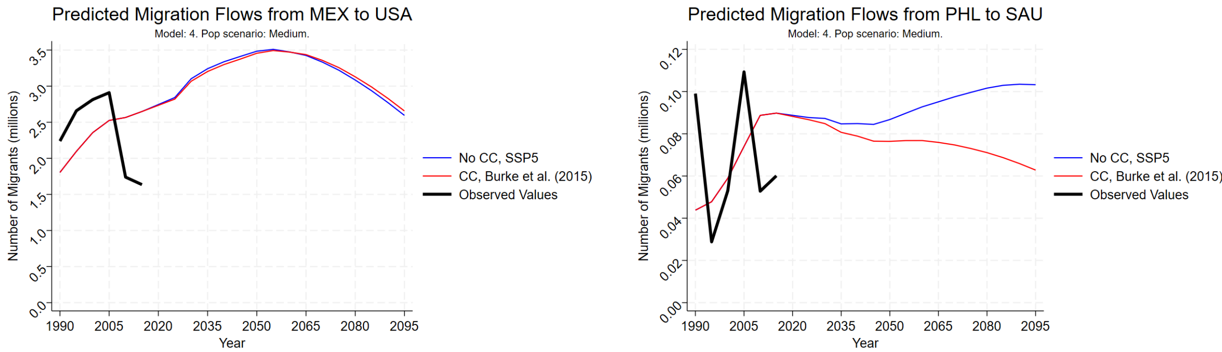
However, the exercise is rich and heterogeneous enough that it is easy to find particular corridors where the two scenarios lead to drastically different migration flows. Figure 10 presents two examples in two relevant corridors. To the left, we can see that Mexican flows to the United States would hardly be affected by changes in GDP associated to climate change. To the right, however, climate change is predicted in this model to severely decrease migration flows from the Philippines to Saudi Arabia.

5.5 Uncertainty

Even if all of the prediction exercises presented so far allow us to see that a variety of scenarios could be justified with very simple gravity models, it must be recalled that each of these scenarios would actually carry some degree of uncertainty.

In order to understand the degree of uncertainty associated to each of the scenarios, we show next the standard errors that we associate to our baseline model in the case of the

Figure 10: Examples of Small and Large Effects of Climate Change on Migration Flows



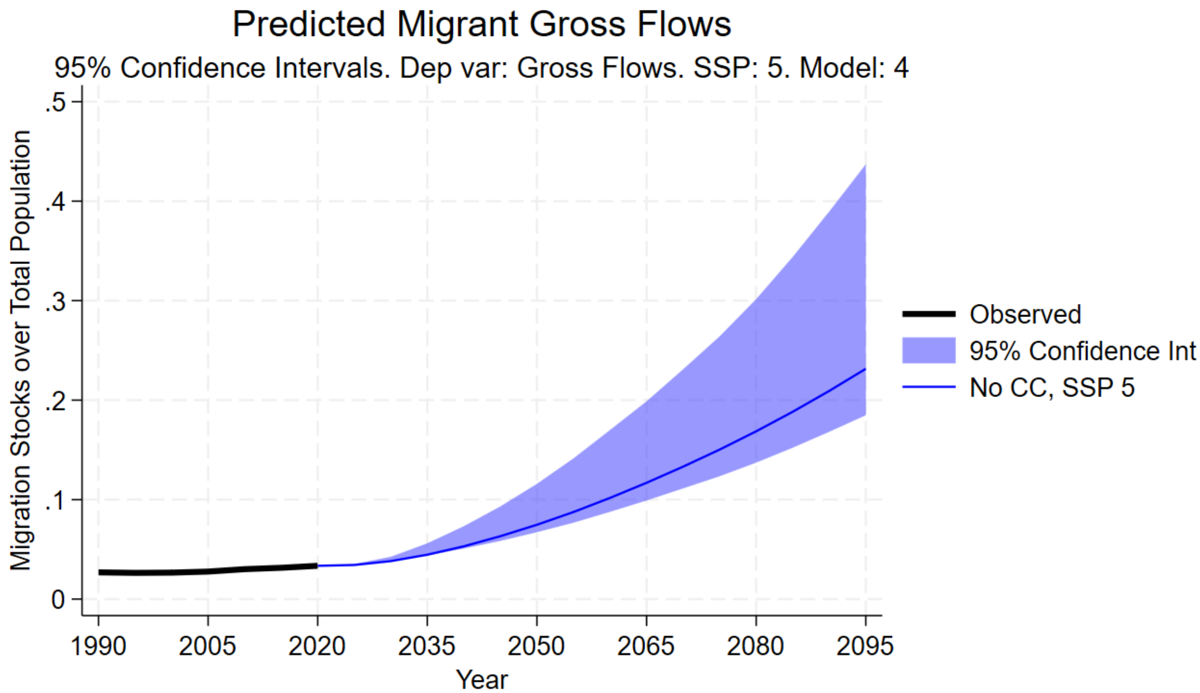
Source: own elaboration. CC refers to climate change, defined as the baseline decrease in GDP projected by Burke et al. (2015). The left panel shows predicted migration flows from Mexico to the United States. The right panel shows predicted migration flows from the Philippines to Saudi Arabia.

medium population variant and an evolution of GDP according to SSP5. To calculate these standard errors, given that we needed to update population and migration stocks for every year, what we have done is to bootstrap the whole procedure, sampling with replacement 1,000 times out of our baseline database.

Figure 11 shows the results of the exercise. The conclusion with respect to the uncertainty of the models is very similar to the one that can be reached by looking at different scenarios by population and economic evolution of the countries. In the medium term, roughly until 2050, the uncertainty is not that large, with 95 per cent of the replications falling between 6.5 and 11.5 per cent of the world population as a likely world migration stock. However, in the long term, the uncertainty skyrockets, even within a simple model out of the many scenarios that have been pointed at. By 2095, the 95 per cent confidence interval for the world migration stock spans from between 20 and 45 per cent of the world population.

We need to emphasize again that this uncertainty would be associated to one single scenario All in all, we have hinted so far in this section to 1,500 scenarios, the result of 5 specifications in terms of variables (Figure 5), 5 in terms of dependent variables (Figure 6), 10 in terms of population variants (Figure 7) and 6 in terms of GDP scenarios (Figure 8).

Figure 11: Predicted Share of Migrants with Confidence Interval



Source: own elaboration. 95 per cent confidence interval based on 1,000 bootstrapped replications with replacement.

5.6 Comparison with Welch and Raftery (2022)

Given the large number of scenarios and confidence intervals emphasized above, it is reasonable to wonder to what extent the gravity model presented here can be useful as a prediction tool. This is why we end this section by comparing the predictive performance of our baseline model with that in Welch and Raftery (2022), arguably one of the best prediction models the academic literature can offer.

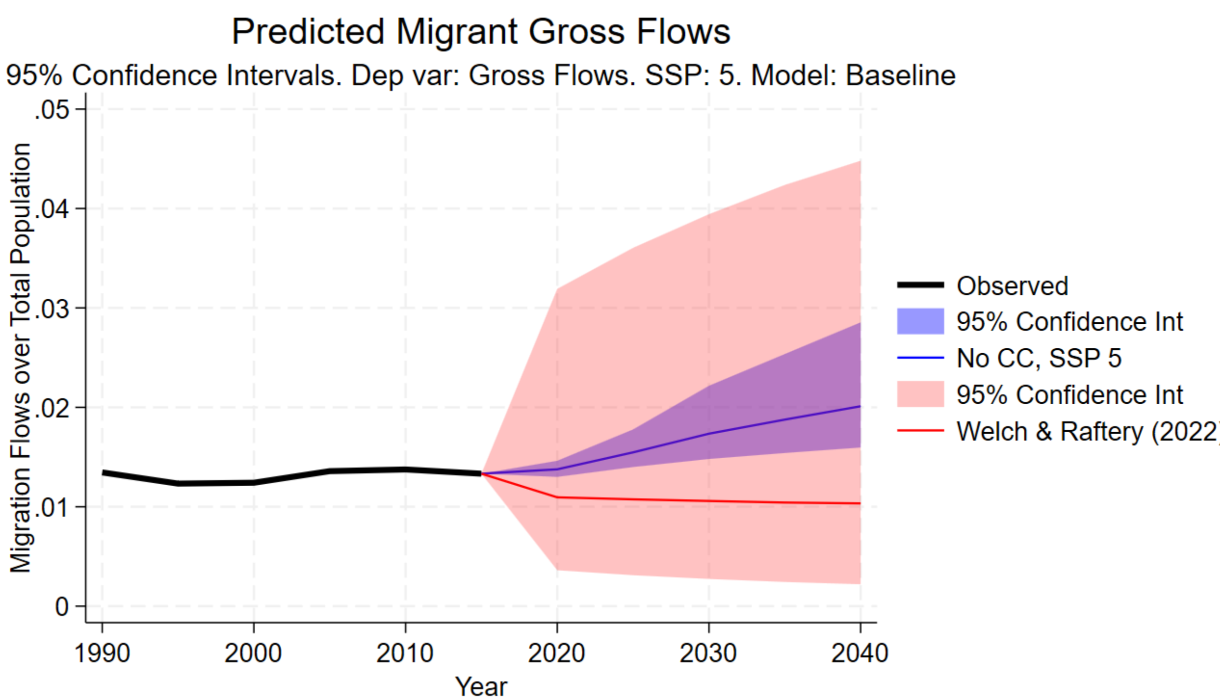
Welch and Raftery (2022) justify the superiority of their Bayesian approach to forecasting migration flows by fitting it to 1990-2015 data and then using it to predict 2015-2020 flows. The mean absolute error of their prediction is 1.2 thousand migrants. Our baseline model reaches 1.4,[16] although our model on net flows goes down to 0.9. We are certainly below the 3.0 thousand that they report for their “Gravity Model” and the 10.0 thousand of their “Poisson hurdle model.”

[16] Our baseline model up to 2015 is presented in Appendix A.3.

Our model does worse in terms of mean absolute prediction error, though. While Welch and Raftery (2022) reach 76 per cent, our baseline model stays at 164 per cent. It is still much lower than their reported "Gravity Model" in Welch and Raftery (2022), at 1,565 per cent, and than their "Poisson hurdle model," which goes up to 25,649 per cent.

However, one advantage of our baseline gravity model is the size of the confidence interval associated to it. Figure 12 compares the 95 per cent confidence interval reported by Welch and Raftery (2022) with the one we obtained following the procedure described above to build Figure 11.

Figure 12: Predicted Migration Flows: our baseline and Welch and Raftery (2022)



Source: own elaboration and Welch and Raftery (2022). Our 95 per cent confidence interval is based on 1,000 bootstrapped replications with replacement.

Figure 12 is different from other figures in this paper because it reports migration flows as a share of the population instead of migration stocks. The confidence interval for Welch and Raftery (2022) goes almost from 0 to 4.5 per cent of the world population for the prediction of migration flows between 2040 and 2045. Compared to theirs, ours is much more concentrated between 1.5 and a bit more than 2.5 per cent of the world population of the same period.

Notice that both confidence intervals are asymmetric, as migration flows are bounded below by zero. If we only look at the central prediction, ours is larger than the one in Welch and Raftery (2022), probably because they neglect migration networks in their model.

Still, we acknowledge that this higher level of precision of our model can be considered artificial, given our previous discussion of the many alternative scenarios that we could build with the gravity models, potentially as reasonable as our baseline one, and each of them with its own level of uncertainty.

Finally, Welch and Raftery (2022) also assess the predictive power of their model by computing the R^2 of a simple regression of real bilateral flows on their predicted flows for 2015-2020. They report an R^2 of 0.97. Our very simple, but much easier to interpret, baseline model features an R^2 of 0.96, again not very far from theirs.

6 Conclusion

This paper has shown the usefulness of the gravity model as a prediction tool for international migration flows. From a theoretical perspective, we have reviewed the arguments that call for the estimation of Poisson Pseudo Maximum Likelihood models with heterogeneous elasticities by origin and destination. PPML models fit migration flows directly while OLS models on ratios are subject to the retransformation problem and typically lead to underestimate the prediction of flows.

Our most novel result has to do with the relevance of network effects, defined as the stock of co-nationals already residing in a destination countries, for the prediction of migration flows. Models estimated on net migration flows do not find a correlation with networks and tend to predict much fewer migration flows into the future than models estimated on gross migration flows.

Consistent with the earlier literature, we have also shown that gravity predictions respond to the evolution of demographic conditions at origin, particularly the size of the population in origin countries, but they are not very sensitive to the evolution of demographic conditions at destination or economic variables both at origin and at destination. For example, the elasticity of migration flows with respect to GDP per capita is typically small and only significant with respect to the GDP per capita at destination.

This feature leads to generally consistent predictions across models and demographic and socioeconomic scenarios in the medium term. In the longer term, however, we find both a much higher uncertainty and a wild variation in scenarios.

We believe the main advantage of the gravity model as a prediction tool is its transparency. The gravity model provides us with a tool that allows us clearly to see the impact of different assumptions of the potential path of migration flows or, more precisely, migration pressures, over the near future. We can see to what extent different assumptions on demography, the role of networks or economic conditions map into different paths for future international migration. The cost in terms of predictive ability with respect to other leading models in the literature is minimal. Paraphrasing the famous misquote attributed to Mark Twain, the reports of the death of the gravity model as a prediction tool for migration flows have been greatly exaggerated.

References

Abel, G. (2022): "Bilateral international migration flow estimates," .

Abel, G. J. and J. E. Cohen (2019): "Bilateral international migration flow estimates for 200 countries," *Scientific Data*, 6, 82.

Abel, G. J. and N. Sander (2014): "Quantifying Global International Migration Flows," *Science*, 343, 1520–1522.

Azevedo, J. P. (2011): "WBOPENDATA: Stata module to access World Bank databases," Statistical Software Components, Boston College Department of Economics.

Azose, J. J. and A. E. Raftery (2019): "Estimation of emigration, return migration, and transit migration between all pairs of countries," *Proceedings of the National Academy of Sciences*, 116, 116–122.

Backhaus, A., I. Martinez-Zarzoso, and C. Muris (2015): "Do climate variations explain bilateral migration? A gravity model analysis," *IZA Journal of Migration*, 4, 3.

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

Beine, M., M. Bierlaire, E. Paschalidis, S. Varotto, and A. B. Vortisch (2024): "The Impact of a Possible Trump Reelection on Mexican Immigration Pressures in Alternative Countries," Tech. rep.

Beine, M., F. Docquier, and c. Özden (2011): "Diasporas," *Journal of Development Economics*, 95, 30–41.

Beine, M. and C. Parsons (2015): "Climatic Factors as Determinants of International Migration," *The Scandinavian Journal of Economics*, 117, 723–767.

Benveniste, H., J. C. Cuaresma, M. Gidden, and R. Muttarak (2021): "Tracing international migration in projections of income and inequality across the Shared Socioeconomic Pathways," *Climatic Change*, 166, 39.

Bertoli, S. (2017): "Is the Mediterranean the New Rio Grande? A Comment," *Italian Economic Journal: A Continuation of Rivista Italiana degli Economisti and Giornale degli Economisti*, 3, 255–259.

Bertoli, S., F. Docquier, H. Rapoport, and I. Ruyssen (2021): "Weather shocks and migration intentions in Western Africa: insights from a multilevel analysis," *Journal of Economic Geography*, 22, 289–323.

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

——— (2015): "The size of the cliff at the border," *Regional Science and Urban Economics*, 51, 1–6.

Beyer, R. M., J. Schewe, and H. Lotze-Campen (2022): "Gravity models do not explain, and cannot predict, international migration dynamics," *Humanities and Social Sciences Communications*, 9, 56.

Borjas, G. J. (2003): "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration in the Labor Market," *Quarterly Journal of Economics*, 118, 1335–1374.

Boss, K., A. Groeger, T. Heidland, F. Krueger, and C. Zheng (2024): "Forecasting bilateral asylum seeker flows with high-dimensional data and machine learning techniques," *Journal of Economic Geography*, lbae023.

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

Burzyński, M., C. Deuster, F. Docquier, and J. de Melo (2022): "Climate Change, Inequality, and Human Migration," *Journal of the European Economic Association*, 20, 1145–1197.

Burzyński, M., F. Docquier, and H. Scheewel (2021): "The geography of climate migration," *Journal of Demographic Economics*, 87, 345–381.

Böhme, M. H., A. Gröger, and T. Stöhr (2020): "Searching for a better life: Predicting international migration with online search keywords," *Journal of Development Economics*, 142, 102347, special Issue on papers from "10th AFD-World Bank Development Conference held at CERDI, Clermont-Ferrand, on June 30 - July 1, 2017".

Campos, R. G. (2017): "International migration pressures in the long run," Working Papers 1734, Banco de España.

Cattaneo, C. and G. Peri (2016): "The migration response to increasing temperatures," *Journal of Development Economics*, 122, 127–146.

Clemens, M. A. (2014): "Does development reduce migration?" in *International Handbook on Migration and Economic Development*, Edward Elgar Publishing, Chapters, chap. 6, 152–185.

Clemens, M. A. and M. Mendola (2024): "Migration from developing countries: Selection, income elasticity, and Simpson's paradox," *Journal of Development Economics*, 171, 103359.

Conte, M., P. Cotterlaz, and T. Mayer (2022): "The CEPII Gravity Database," Working Papers 2022-05, CEPII research center.

Correia, S., P. Guimarães, and T. Zylkin (2020): "Fast Poisson estimation with high-dimensional fixed effects," *The Stata Journal: Promoting communications on statistics and Stata*, 20, 95–115.

Dao, T. H., F. Docquier, M. Maurel, and P. Schaus (2021): "Global migration in the twentieth and twenty-first centuries: the unstoppable force of demography," *Review of World Economics*, 157, 417–449.

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

Desmet, K. and E. Rossi-Hansberg (2015): "On the spatial economic impact of global warming," *Journal of Urban Economics*, 88, 16–37.

Docquier, F. and J. Machado (2017): "Income Disparities, Population and Migration Flows Over the Twenty First Century," *Italian Economic Journal*, 3, 125–149.

Docquier, F., c. Özden, and G. Peri (2014): "The Labour Market Effects of Immigration and Emigration in OECD Countries," *The Economic Journal*, 124, 1106–1145.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015): "The Next Generation of the Penn World Table," *American Economic Review*, 105, 3150–3182.

Fernández-Huertas Moraga, J. and G. López Molina (2018): "Predicting Spanish Emigration and Immigration," Working Papers 2018/6, AIReF.

Gidden, M. J., K. Riahi, S. J. Smith, S. Fujimori, G. Luderer, E. Kriegler, D. P. van Vuuren, M. van den Berg, L. Feng, D. Klein, K. Calvin, J. C. Doelman, S. Frank, O. Fricko, M. Harmsen, T. Hasegawa, P. Havlik, J. Hilaire, R. Hoesly, J. Horing, A. Popp, E. Stehfest, and K. Takahashi (2019): "Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century," *Geoscientific Model Development*, 12, 1443–1475.

Hanson, G. H. and C. McIntosh (2010): "The Great Mexican Emigration," *The Review of Economics and Statistics*, 92, 798–810.

——— (2012): "Birth Rates and Border Crossings: Latin American Migration to the US, Canada, Spain and the UK," *Economic Journal*, 122, 707–726.

——— (2016): "Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run," *Journal of Economic Perspectives*, 30, 57–82.

Kluge, L., S. KC, D. Yildiz, G. Abel, J. Schewe, and O. Olaya-Bucaro (2024): "A multidimensional global migration model for use in cohort-component population projections," *Demographic Research*, 51, 323–376.

Manning, W. G. (1998): "The logged dependent variable, heteroscedasticity, and the retransformation problem," *Journal of Health Economics*, 17, 283–295.

McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

Mckenzie, D. and H. Rapoport (2007): "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics*, 84, 1–24.

Munshi, K. (2003): "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market," *The Quarterly Journal of Economics*, 118, 549–599.

Ottaviano, G. I. P. and G. Peri (2012): "Rethinking The Effect Of Immigration On Wages," *Journal of the European Economic Association*, 10, 152–197.

Özden, c., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011): "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000," *World Bank Economic Review*, 25, 12–56.

Raymer, J., A. Wiśniowski, J. J. Forster, P. W. F. Smith, and J. Bijak (2013): "Integrated Modeling of European Migration," *Journal of the American Statistical Association*, 108, 801–819.

Riahi, K., D. P. van Vuuren, E. Kriegler, J. Edmonds, B. C. O'Neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, A. Popp, J. C. Cuaresma, S. KC, M. Leimbach, L. Jiang, T. Kram, S. Rao, J. Emmerling,

K. Ebi, T. Hasegawa, P. Havlik, F. Humpenöder, L. A. D. Silva, S. Smith, E. Stehfest, V. Bosetti, J. Eom, D. Gernaat, T. Masui, J. Rogelj, J. Strefler, L. Drouet, V. Krey, G. Luderer, M. Harmsen, K. Takahashi, L. Baumstark, J. C. Doelman, M. Kainuma, Z. Klimont, G. Marangoni, H. Lotze-Campen, M. Obersteiner, A. Tabeau, and M. Tavoni (2017): "The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview," *Global Environmental Change*, 42, 153–168.

Rogelj, J., A. Popp, K. V. Calvin, G. Luderer, J. Emmerling, D. Gernaat, S. Fujimori, J. Strefler, T. Hasegawa, G. Marangoni, V. Krey, E. Kriegler, K. Riahi, D. P. van Vuuren, J. Doelman, L. Drouet, J. Edmonds, O. Fricko, M. Harmsen, P. Havlík, F. Humpenöder, E. Stehfest, and M. Tavoni (2018): "Scenarios towards limiting global mean temperature increase below 1.5 °C," *Nature Climate Change*, 8, 325–332.

Rustemeier, E., S. Hänsel, P. Finger, U. Schneider, and M. Ziese (2022): *GPCC Climatology Version 2022 at 0.5°: Monthly Land-Surface Precipitation Climatology for Every Month and the Total Year from Rain-Gauges built on GTS-based and Historical Data*, `10.5676/DWD_GPCC/CLIM_M_V2022_050`.

Santos Silva, J. M. C. and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88, 641–658.

Sjaastad, L. A. (1962): "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70, 80–93.

United Nations (2020): "International Migrant Stock 2020." Department of Economic and Social Affairs, Population Division. (United Nations database, POP/DB/MIG/Stock/Rev.2020).

——— (2022): "World Population Prospects 2022," Department of economic and social affairs, population division. online edition.

University of Delaware (2018): *Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 2017). Version 5.01*, `https://www.esrl.noaa.gov/psd/data/gridded/data.UDel_AirT_Precip.html`.

Welch, N. G. and A. E. Raftery (2022): "Probabilistic forecasts of international bilateral migration flows," *Proceedings of the National Academy of Sciences*, 119, e2203822119.

Willmott, C. J. and K. Matsuura (1995): "Smart Interpolation of Annually Averaged Air Temperature in the United States," *Journal of Applied Meteorology*, 34, 2577–2586.

World Bank (2024): *World Development Indicators*, `http://databank.worldbank.org`.

A Appendix

A.1 Sample selection

We start from the original data on migration flows by Abel (2022), disaggregated by gender. After adding male and female migrants by corridor, we are left with 304,658 observations. We then aggregate Sudan and South Sudan and also Serbia and Montenegro, the two countries that split during the sample period so that the panel of countries does not change. This leaves us with 304,654 observations, losing 0.28 per cent of the gross international migration flows that become internal migration due to the aggregation.

When we merge with population data from United Nations (2022), the only observations we do not match come from the population side, hence we do not lose migration observations.

Next, for the network variable, the stock of migrants from the same origin at the same destination at the beginning of the period, we go back to United Nations (2020). Despite the fact that Abel (2022) is built from United Nations (2020), we fail to match many corridors. Luckily, their quantitative relevance is generally very small. From the stock side, we do not match 728 observations, but they amount to less than 8,000 migrants globally per five-year period. For example, the largest missing corridor is 4,095 immigrants from the Caribbean Netherlands (Bonaire, Sint Eustatius and Saba) living in Canada in 2015. A larger concern is our failure to match 70,342 observations for which we have flow data but no stock data. This is 77 per cent of the sample in terms of observations, but only 5.8 per cent in terms of gross migration flows. Out of the distribution of gross migration flows corresponding to these missing migration stocks, 73 per cent are zero and the 99$^{\text{th}}$ percentile is 1,477 migrants. As a result, we decided to assume these missing corridors were zero so as not to lose the observations. However, this might not be true for some corridors. For example, we are assuming a zero migration stock for a gross migration flow of 1,415,483 migrants going from Iran to Afghanistan between 1990 and 1995. The estimates in the paper are robust to just dropping all these non-matched observations or to assuming heterogeneous coefficients for these corridors.[17]

Then, we merge the GDP data from the Penn World Tables (Feenstra et al., 2015) and from the World Bank World Development indicators (World Bank, 2024). We fail to match 73,330 observations, which amount to 5 per cent of gross migration flows. This leaves us

[17]Results available from the authors upon request.

with a total number of observations of 231,324 in our baseline sample. Table 6 provides basic summary statistics on this baseline sample.

Table 6: Summary Statistics

	Obs	Mean	SD	Min	Max
Dependent Variables: Migration					
Gross Flows (Pseudo-Bayesian)	231,324	2,095	26,482	0	2,909,903
Net Flows (Stock Diff Drop Negatives)	231,324	706	15,425	0	2,709,438
Net Flows (Stock Diff Rev Negatives)	231,324	889	16,997	0	2,709,438
Gross Flows (Open Accounting)	231,324	870	16,819	0	2,786,012
Gross Flows (Closed Accounting)	231,324	1,089	18,118	0	2,217,482
Independent Variables					
Population at Destination	231,324	31,571,298	124,136,741	4,673	1,389,794,304
Share 15-49 at Destination	231,324	0.506	0.050	0.400	0.857
Migrant Stock	231,324	4,515	76,349	0	12,168,662
GDP per Capita at Destination	231,324	16,486	19,089	266	129,818
Distance	219,376	8,120.444	4,546.672	8.000	19,819
Contiguous	219,376	0.015	0.121	0	1
Common Language	219,376	0.172	0.378	0	1
Former Colony	219,376	0.010	0.097	0	1
Average Temperature at Destination	167,682	18.797	8.071	-8.885	29
Total Precipitation at Destination	167,682	1,112	816	23	4340

Migration flows, stocks and population measured in number of persons, distance is measured in kilometers, GDP per capita in PPP dollars, temperatures are measured in Celsius degrees and precipitations in millimeters.

When we merge our baseline with climate data, missing temperature data for small territories lead us to lose 63,642 additional observations, which account for 3.5 per cent of our baseline gross migration flows.

A.2 Prediction Models for Figure 5

Table 7: Prediction models for Figure 5: gross flows

	(1)	(2)	(3)	(4)	(5)
Log Population at Origin	1.360***	1.420***	2.145***	2.097***	1.918***
	(0.275)	(0.271)	(0.285)	(0.285)	(0.266)
Share 15-49 at Origin		-0.843	-1.960*	-1.872	-1.420
		(1.161)	(1.147)	(1.216)	(1.147)
Log Population at Destination			-0.887***	-1.178***	-1.118***
			(0.228)	(0.201)	(0.269)
Share 15-49 at Destination			2.651***	3.481***	3.085***
			(0.904)	(0.846)	(0.962)
Log (Migrant Stock +1)				0.152***	0.143***
				(0.048)	(0.051)
Log GDPpcPPP at Destination					0.196**
					(0.096)
Log GDPpcPPP at Origin					-0.093
					(0.152)
Observations	231324	231324	231324	231324	231324
Pseudo R^2	0.953	0.953	0.954	0.955	0.955
Dyadic FE	Yes	Yes	Yes	Yes	Yes

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors by origin country in parentheses. The dependent variable are pseudo-Bayesian estimates of gross flows following Azose and Raftery (2019), provided by Abel (2022). Dyadic FE refers to origin-destination fixed effects.

A.3 Baseline model up to 2015

Table 8: Baseline estimation on 1990-2015 flows

Log Population at Destination	-1.417***
	(0.314)
Log Population at Origin	2.243***
	(0.335)
Share 15-49 at Destination	4.563***
	(1.068)
Share 15-49 at Origin	-2.024
	(1.239)
Log (Migrant Stock +1)	0.178***
	(0.045)
Log GDPpcPPP at Destination	0.102
	(0.072)
Log GDPpcPPP at Origin	0.053
	(0.123)
Observations	191524
Dyadic FE	Yes
Pseudo R^2	0.960

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors by origin country in parentheses. The dependent variable are pseudo-Bayesian estimates of gross flows following Azose and Raftery (2019), provided by Abel (2022). Dyadic FE refers to origin-destination fixed effects.