

## **CSAE Working Paper WPS/2014-27**

### **Migration, Diasporas and Culture: an Empirical Investigation**

Paul Collier\* and Anke Hoeffler\*\*

August, 2014

#### **Abstract**

Using global data we examine the dynamics of migration from developing to developed countries. Origin and destination countries are characterized by substantial differences in incomes, political rights and cultures. Incentives as well as costs shape the decision to migrate. One powerful dynamic effect is that diasporas increase migration, mainly because they lower the cost of migration. Diasporas assist the next wave of migrants by overcoming the high cost of the emigration, in particular when the origin country is far away and poor. The interaction between the diaspora and cultural distance is also significant. Diasporas in culturally distant countries appear to be particularly useful in overcoming the cost of migration. Culturally distant diasporas are less likely to assimilate and maintain closer links with their country of origin, while diasporas from culturally similar countries are more likely to assimilate and thus be less useful to potential new migrants.

**Keywords:** Migration, development, culture

**JEL Codes:** O15, Z1

\*Blavatnik School of Government Oxford University, and Centre for the Study of African Economies

\*\* Centre for the Study of African Economies, Department of Economics, Oxford University. Manor Road Building, Manor Road, Oxford OX1 3UQ, [anke.hoeffler@economics.ox.ac.uk](mailto:anke.hoeffler@economics.ox.ac.uk)

We would like to thank Michel Beine, Simone Bertoli, Stephen Fisher and Mariola Pytlikova for comments on an earlier draft and David Hendry and Jennifer Castle for help with Oxmetrics. All remaining errors are our own.

Anke Hoeffler's participation was funded through the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 290752.



**Centre for the Study of African Economies**

Department of Economics · University of Oxford · Manor Road Building · Oxford OX1 3UQ

**T:** +44 (0)1865 271084 · **F:** +44 (0)1865 281447 · **E:** [csae.enquiries@economics.ox.ac.uk](mailto:csae.enquiries@economics.ox.ac.uk) · **W:** [www.csae.ox.ac.uk](http://www.csae.ox.ac.uk)

## Highlights

- We model migration flows from low and middle income to high income countries.
- Incentives and costs shape the decision to migrate.
- The existing stock of migrants, the diaspora, helps to overcome the cost of migration.
- Diasporas in culturally distant countries are particularly useful in overcoming the cost of migration.
- Diasporas from culturally similar countries are less useful to potential new migrants, because they are more likely to assimilate in their host country.

# Migration, Diasporas and Culture: an Empirical Investigation

## 1. Introduction

The increase in international migration since 1945 is often considered as an integral part of globalization. However, this is belied by its composition. Migration among OECD countries has been broadly constant despite a vast expansion in both trade and capital flows between them. The expansion in global migration is entirely accounted for by flows from developing countries to the OECD. Developed countries account for 15 per cent of the global population, but are the destination of 56 per cent of all migrants. While the defining difference between developed and developing countries is income, there are also substantial differences in political freedom and in culture. Migration from developing to developed countries is likely to be driven by these three gaps and their interactions.

In this paper we analyze the evolution of the flow of voluntary migration from developing to developed countries. We focus on voluntary migration by excluding movements of refugees. Using recently released global panel data, we are able to estimate inter-decade flows for each decade between 1960 and 2000. While there is a rich literature on migration to OECD countries (for a survey see Docquier and Rapoport, 2012) the studies typically use only one or two waves of data.

We show that the evidence supports a conceptualization of the act of migration as an investment on which the migrant gets a subsequent return. The return reflects the gaps in income and political freedom between their host and home countries, and the required investment reflects a range of impediments to mobility. Potentially, one important impediment is a difference in cultures: wide culture gaps are as evident as income gaps. The investment costs of migration are reduced by support from prior migrants who have retained their connections to their home societies. Consistent with Beine *et al.* (2011a), we find that this effect of diasporas is powerful. A neglected implication is that migration has a tendency to accelerate: migration builds diasporas, and diasporas ease subsequent migration. However, this role of diasporas introduces a further potential effect of culture. The cultural distance between home and host societies may affect the rate at which immigrants absorb into their host society, and hence the rate at which their connections with their country-of-origin decay. If a wide culture gap preserved the link between immigrants and their home societies, it would increase the gearing from the stock of immigrants onto the subsequent flow. A large difference in cultures could therefore potentially both impede migration and accelerate its

increase. This paper is the first to estimate these effects of income, politics and culture in an integrated approach.

In Section 2 we estimate a model of migration, using a set of economic and political explanatory variables, and show how the results support an investment-return interpretation. We introduce a measure of cultural distance in Section 3 and show that it has powerful and somewhat counter-intuitive effects. Section 4 concludes.

## 2. Estimating a Core Model

### *The Dependent Variable*

In our model we analyse the flow of migrants from low and middle income countries to high income OECD countries over the period 1970-2000. The classification is based on the World Bank's country definition for 2000, the most recent year for which comprehensive migration data are available. As we explain further below, migrants are defined as people who were not born in their current country of residence. This information is available in country pair format. These country pairs are directed dyads,  $m_{ij}$ , where  $i$  denotes the country of origin and  $j$  the destination country. These dyads are directed because the migration stock for country pair  $ij$  is not the same as for country pair  $ji$ . Since we only consider cross-border migration we do not consider any movements within countries and do not analyse  $m_{ii}$ ,  $m_{jj}$  etc. The total number of country pairs is therefore  $N(N - 1)$ .

In our migration model we model the change in the stock of migrants, or the flow of migrants, over each decade:

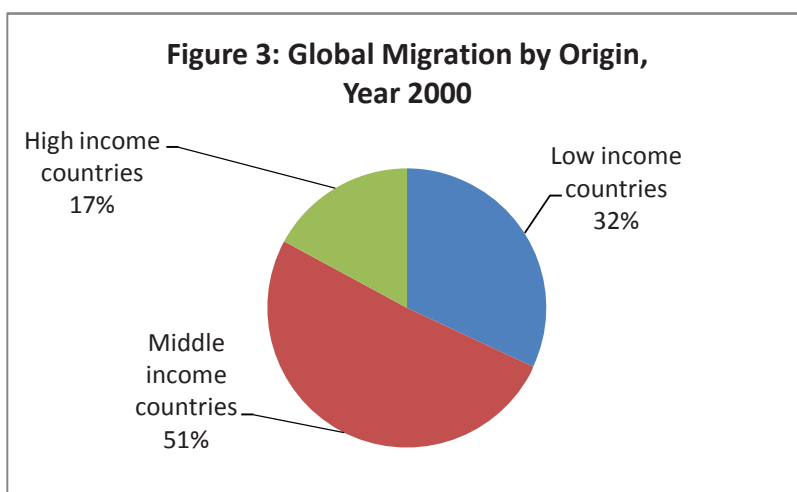
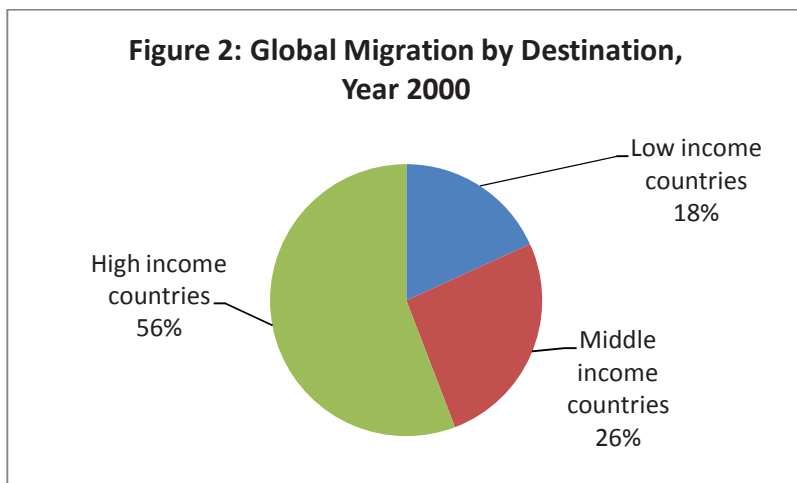
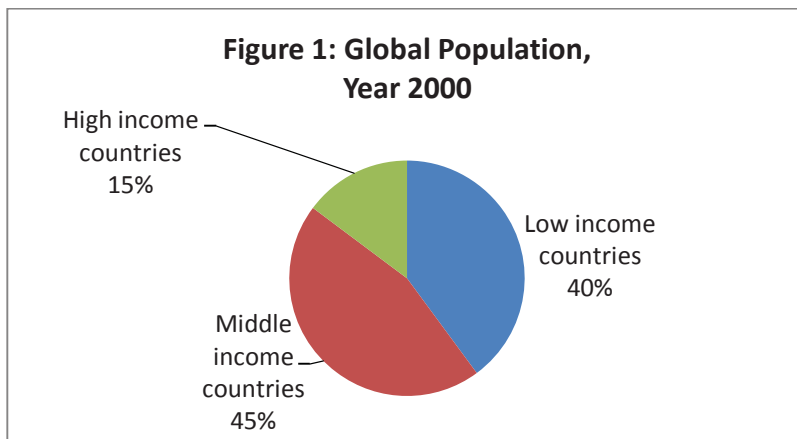
$$m_{ijt} - m_{ijt-1} = \alpha + \beta_0 m_{ijt-1} + \beta_1 z_{it} + \beta_2 z_{jt} + \beta_3 w_{ij} + \beta_4 m_{ijt-1} \cdot z_{it} + \beta_5 m_{ijt-1} \cdot z_{jt} + \beta_6 m_{ijt-1} \cdot w_{ij} + \eta_i + \eta_j + \tau_t + e_{ijt} \quad (1)$$

where  $i$  and  $j$  are the origin and destination country, respectively and  $t$  denotes time (1960, 1970, 1980, 1990 and 2000). The stock of migrants at time  $t$  is denoted by  $m_{ijt}$  and the difference between the stock of migrants at time  $t$  and  $t-1$  approximates the flow of migrants ( $m_{ijt} - m_{ijt-1}$ ). Our dyadic dataset contains two types of explanatory variables. One set is specific to the country, be it the country of origin or destination, for example income ( $z_{it}$ ,  $z_{jt}$ ). Another set of variables is specific to the country pair, such as a common border or colonial

history ( $w_{ij}$ ). We expect that the effect of the country characteristics depends on the size of the existing migrant stock, or the diaspora. We therefore interact the country and country pair characteristics with the diaspora (e.g.  $w_{ij} \cdot m_{ijt-1}$ ). We also include country fixed effects for the origin countries as well as destination countries,  $\eta_i$  and  $\eta_j$ , and a time effect,  $\tau_t$ . We also note that the error terms  $e_{ij}$  are likely to be correlated across observations.

The data used for this paper come from Özden *et al* (2011). They define migrants as people living in one country but having been born in another country. Thus, the concept of migration is based on where people are born and not on their nationality. The data set covers all countries and provides the number of migrants originating from country  $i$  now living in country  $j$ . The data set provides migrant population numbers for 1960, 1970, 1980, 1990 and 2000 in 175 countries.

The following figures and tables provide some description of the data. Figures 1-3 show that while the high income OECD countries only account for 15 per cent of the global population, they are the destination of 56 per cent of all global migrants (originating from poor, middle and high income countries).



**Source:** Özden *et al* (2011), Penn World Tables, own calculations. The country classifications are based on the World Development Report 2000/2001.

Table 1 provides a further breakdown of the origin and destination of migrants. In 2000 about 70 million migrants from poor and middle income countries were living in the high income OECD countries. This accounts for about 42 per cent of all global migration. In our analysis we concentrate on the examination of the migrant flows from poor and middle income countries to high income OECD countries.

**Table 1: Stock of Migrants in the Year 2000**

Origin down/destination across	To Low income countries	To Middle income countries	To High income countries	Total Migrants (% of total population)	Total population
From Low income countries	21.2	17.9	13.6	52.7 (2.2%)	2380
From Middle income countries	8.19	19.7	56.1	83.99 (3.1%)	2710
From High income countries	0.71	5.09	22.4	28.2 (3.2%)	881
Total	30.1	42.69	92.1	164.89 (2.8%)	5971

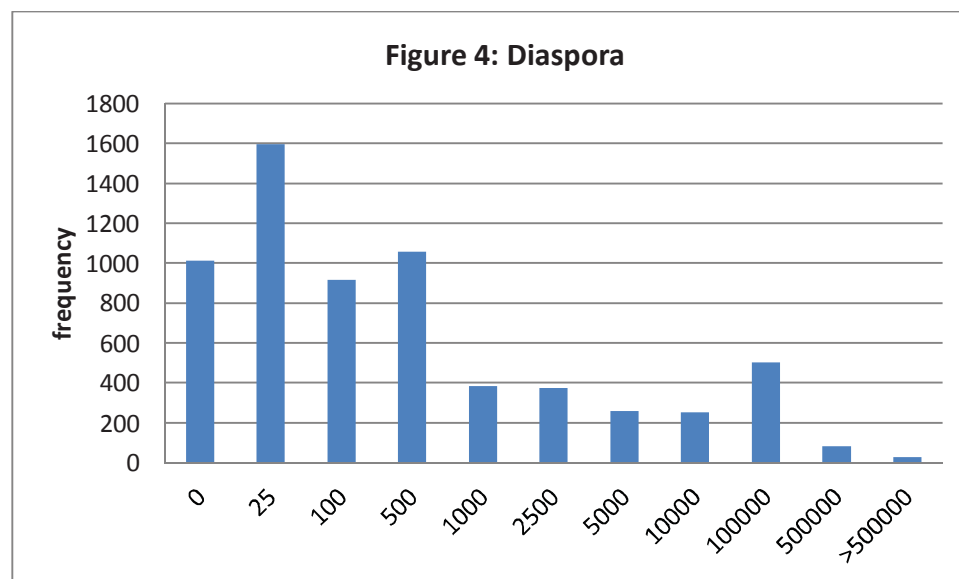
Note: Numbers in millions Source: Özden *et al* (2011), Penn World Tables, own calculations. The country classifications are based on the World Development Report 2000/2001.

Özden *et al* (2011) provide the first data set to cover global migration over time. While the data coverage is impressive, the use of these data poses some statistical challenges. We only highlight two issues here; Beine *et al* (2011b, 2014) discuss these and other challenges in more detail.

#### *‘Zero Observations’*

One characteristic of the migration data is that many dyads take a value of zero, because nobody born in country  $i$  now lives in country  $j$ . OLS may therefore not be the appropriate estimation method. In our regression sample we have 6379 observations of which 682 are

zero.<sup>1</sup> One option would be to disregard the zeros and only use the positive observations for our analysis. However, it is not clear whether many of the zeros are indeed zeros or are in fact small numbers. In many countries published census data does not provide information on very small groups of people, because they could be identified. Zero or small stocks of migrants that did not scale up over time are an interesting phenomenon and we would like to consider this in our analysis, thus, we decided to keep them in our sample. Figure 4 provides a frequency distribution of the stock of migrants, to which we also refer as the diaspora.



### *Stocks and Flows*

Ideally we would like to analyse the determinants of the inflow of new migrants. However, we only have data on the stock of migrants at five points in time. The difference in the stocks between time  $t-1$  and  $t$  is not equal to the inflow of new migrants because some of the migrants present at time  $t-1$  will have died, returned or migrated to a different country by time  $t$ . Thus, the difference between the stocks of migrants between time  $t-1$  and  $t$  will understate the inflow of new migrants. We do not have sufficient information on return migration but we do have some information on mortality. How many migrants survive from one round of censuses to the next depends on the age structure of the migrant population and the mortality rates of these migrants in their host country. In the Appendix we outline how we used data on the age structure of the migrant stock and host country mortality rates to adjust

<sup>1</sup> We investigated the possibility of sample selection bias by estimating a Heckman model. As it is often the case with Heckman estimations, it is difficult to think of a variable that determines whether or not there is any migration for country pair  $ij$  but not how many people migrate. As suggested by Beine *et al* (2011b) we use the existence of diplomatic relations in 1960 as the selection variable, but found the inverse Mills Ratio to be insignificant. We thus conclude that we can find no evidence for sample selection bias.



the flow of migrants. Basically, we added the estimated number of migrants that died during the decade to the difference of the migrant stocks at the beginning and the end of the decade to derive an adjusted measure of the flow of migrants. To illustrate our adjustment of the migration flows we present some of the migrant flow data for migration from India to the UK in Table 2.

**Table 2: Migration Flows from India to the United Kingdom**

	<b>Unadjusted</b>	<b>Adjusted</b>
<b>1960-69</b>	143,896	164,251
<b>1970-79</b>	63,202	98,509
<b>1980-89</b>	29,747	58,454
<b>1990-99</b>	92,507	112,270

To summarize, we are mindful of the various data problems. We examined the treatment of zero observations but could find no evidence of selection bias. In order to approximate the flows of new migrants we adjust the stocks for mortality. With these data issues in mind we now turn to the estimation of our baseline model.

### *The explanatory variables*

Migration has been examined in many different academic disciplines: history, economics, political science, geography, anthropology and sociology; for recent interdisciplinary reviews see Brettell and Hollifield (2008) and de Haas (2010). The core assumption of economic models is that individuals migrate from low- to high-income regions in order to improve their well-being (Ravenstein's (1885, 1889) 'laws of migration'). Migrants are attracted by better economic, political, demographic and environmental factors in other countries. Gravity-based models have expanded on the idea that migration is a function of spatial disequilibria. For example, Lee (1966) added distance, physical barriers, immigration laws and personal characteristics of the migrant to his migration model. While absolute wage differentials have widened, transport costs have fallen, encouraging migration to increase (Hatton and Williamson, 1998).

We regress the flow of migrants between country dyads on measures of that proxy the incentive to migrate and the cost of migration. The fundamental economic incentive to migrate is the absolute difference in per capita income between the host and home economies. However, while income in the host country has a straightforward incentive effect, that in the country of origin has a dual role since it affects not only the incentive to migrate

but the ability to finance the required investment. Hence, rather than use the income difference as an explanatory variable, we enter incomes in the host and home economies separately. We measure the income variables at the beginning of the decade in order to reduce endogeneity.

While the potential migrant has no choice as to country-of-origin, the country of destination is a choice variable. The migrant determines where to migrate as well as whether to migrate. Bertoli and Fernandez-Huertas Moraga (2013) consider these alternative choices: migration between two countries does not depend solely on their relative attractiveness, but also on the one of alternative destinations. Using high frequency data they account for ‘multilateral resistance’ in their model of migration to Spain. With our large- $n$ , small- $t$  data set we cannot employ their estimation technique but include a relative income variable that captures the alternative choice. In estimating the rate of migration from country  $i$  to  $j$ , where  $j$  is some particular OECD country, we therefore investigate whether the income of  $j$  relative to other potential OECD destinations has significant effects. We again use lagged values.

Since the decision to migrate is intrinsically forward-looking, the pertinent economic variable may be the expectation of future incomes. However, the growth of per capita income in home and host countries over the course of the decade is potentially endogenous to migration during the decade: for example, immigration may raise per capita growth. To control for this we proxy growth of the destination country’s economy by the weighted growth rates of the countries which are the markets for its exports.

Analogous to the economic incentive to migrate, there is an incentive to societies that offer greater political freedom. While there is little variation among OECD societies, there is considerable variation both between developing countries and over time within them. As with income, the extent of political freedom in countries-of-origin has ambiguous effects: some restrictions on freedom may directly or indirectly reduce emigration. Hence, as with income, we enter the degree of political freedom separately for host and home societies. We also include social disruption as proxied by periods of civil and international warfare.

The investment cost of migration is influenced by a range of variables. The most evident proxy variable is the distance between the country dyads. As noted above, diasporas lower the costs of migration. Relatives established in host countries provide finance for travel,

hospitality on arrival, and networks for employment (see Collier (2013), chapter 6 for a discussion). They can also confer the right to migrate through sponsorship, marriage and political influence on regulations. We proxy the size of the diaspora by the stock of immigrants from the country-of-origin in the host society, measured as discussed above. We also include whether there was a former colonial relationship between the dyads. This may influence decisions through the ease of information flows, or more subtly through expectations set during education, as exemplified by the sense of Britain as the ‘home country’ for West Indians described in the writings of V.S. Naipaul.

Migration is subject to a range of regulations imposed by host societies. These changing regulations have yet to be codified into a tractable global dataset. In the absence of such data we can only resort to proxies. One such proxy is indeed whether there was formerly a colonial relationship, since this affects the regulations imposed by host societies. A second proxy is whether the country-of-origin is landlocked. Other than by direct flight, emigration from such countries requires passage through transit countries which are likely to impose their own regulations, so that the migrant is subject to multiple restrictions. Finally, as discussed below, we introduce country fixed effects. These will capture those differences in regulations between OECD societies which are persistent and which are not idiosyncratic as to particular countries-of-origin. The assumption of persistence becomes less reasonable the longer is the time span of the analysis: we take the thirty-year period 1970-2000. Were the controls exogenous to the demand for migration, and fully effective, then they alone would fully explain the volume of migration. Individual choices might affect the composition of migration but would not affect its totals. In practice, migration controls are seldom absolute limits on volumes and are also seldom fully effective. Hence, the demand for migration can be expected to influence volumes.

Finally, we control for time effects by decade dummies, and for the size of country, through the (lagged) populations of host and origin societies. A well-established result in the literature is that small countries, such as islands, have high rates of emigration.

Because migration is a complex decision, subject to a wide array of potential influences, it is valuable to investigate the interactions between variables. However, even though our model is relatively parsimonious in the use of explanatory variables, there are 31 interaction terms. The inclusion of these interaction terms is likely to introduce multicollinearity. In order to

decide which interaction terms to retain in our core model we used the automated selection process as suggested by Johansen and Nielsen (2009) and Doornik (2009). This establishes which interaction terms are significant. The results are presented in Table 3. Column 1 is the General Unrestricted Model (GUM) with all interaction terms included. The more manageable model that resulted from the automated selection process is presented in column 2, which includes 15 interaction effects.

Our model may suffer from omitted variable bias and we investigated whether including country fixed effects changed our results. In order to examine this, we included all fixed effects for countries of origin and destination in our core model and used the automated process to establish which country effects were significant. The resulting model is presented in column 3. It appears that the coefficients are not much changed. Most of the country-of-origin fixed effects are insignificant with the exceptions of Hungary, Korea, Libya, the Philippines and Poland. In contrast, most destination country dummies are significant, reflecting the difference in immigration policies. This interpretation is supported by Koopmans et al (2012), they compare ten West-European countries, concluding that immigrant rights differ significantly across countries and that these differences are persistent. Unfortunately, the use of the Koopmans et al data would reduce our sample substantially and we are thus not able to test the impact of different immigration rights more formally.

**Table 3: Analysis of Migration Flows**

	(1)	(2)	(3)
	All Interaction Terms	Core Model	Core Model plus Country Dummies
Diaspora	4.930 (2.277)**	4.911 (2.127)**	5.192 (2.146)**
Colonial Ties	-436,185.360 (293,278.809)	-524,780.577 (288,615.628)*	-517,111.006 (282,859.500)*
ln Distance (km)	79,893.354 (54,618.567)	55,277.536 (23,047.309)**	75,351.471 (24,424.608)***
Landlocked (dummy)	-1,600.864 (553.864)***	-1,428.186 (524.542)***	249.341 (488.219)
ln Population origin	384.119 (314.446)	413.798 (301.270)	441.954 (280.743)
ln population destination	1,955.285 (437.226)***	1,968.244 (428.298)***	10,446.629 (2,742.483)***
ln GDP origin t-1	4,409.128 (8,325.679)	252.857 (378.978)	1,002.362 (411.665)**
ln GDP destination	64,035.105 (48,844.673)	46,405.190 (23,649.220)**	69,372.786 (27,979.365)**
Growth origin	6,082.405 (5,476.274)	422.381 (222.202)*	466.350 (238.402)*
Growth destination	4,576.086 (22,057.048)	1,259.936 (654.200)*	-3,536.482 (1,408.284)**
Relative GDP	-24,378.344 (62,618.459)	11,270.338 (16,292.803)	12,364.304 (21,061.814)
Polity origin	2,267.423 (2,041.351)	280.284 (72.236)***	308.563 (70.860)***
Polity destination	-294.870 (1,921.462)	228.528 (116.524)**	-145.203 (174.725)
Civil War	1,176.440 (1,132.549)	1,204.098 (1,150.319)	570.184 (1,156.637)
International War	2,213.376 (3,101.173)	2,288.745 (3,027.420)	2,841.200 (2,997.171)
Colonial·ln GDP origin	8,653.564 (3,745.115)**	8,275.607 (3,724.088)**	7,995.639 (3,646.708)**
Colonial·ln GDP destin.	37,027.632 (29,468.520)	47,239.411 (29,118.891)	47,240.864 (28,582.516)*
Colonial·Growth origin	-5,138.955 (3,729.167)	-4,838.823 (3,839.008)	-4,648.869 (3,724.751)
Colonial·Growth destin.	27,877.333 (14,074.792)**	27,272.263 (14,759.390)*	26,606.456 (14,310.639)*
Colonial·Relative GDP	-58,515.970 (40,084.920)	-56,274.398 (39,172.810)	-58,557.623 (38,468.775)
Colonial·Polity origin	-794.001 (543.687)	-778.772 (537.721)	-616.944 (538.178)
Colonial·Polity destin.	1,111.132 (1,208.986)		
ln Distance· lnGDP o	-461.517 (935.026)		
ln Distance· lnGDP d	-8,605.988 (5,532.850)	-6,246.882 (2,436.963)**	-8,640.344 (2,600.825)***
ln dist·Growth origin	-641.344 (609.300)		
ln dist·Growth destin.	-405.030 (2,434.483)		
ln dist·relative GDP	4,501.430 (6,912.348)		
ln dist·Polity origin	-223.682		

---

	(225.770)		
ln dist·Polity destin.	53.232 (214.720)		
Diaspora·ln GDP origin	-0.191 (0.086)**	-0.190 (0.087)**	-0.189 (0.087)**
Diaspora·ln GDP destin.	-0.413 (0.199)**	-0.415 (0.195)**	-0.439 (0.197)**
Diaspora·Growth origin	0.059 (0.103)	0.056 (0.065)	0.047 (0.066)
Diaspora·Growth destin.	-0.009 (0.134)		
Diaspora·relative GDP	0.820 (0.127)***	0.819 (0.129)***	0.829 (0.133)***
Diaspora·Polity origin	0.005 (0.010)	0.005 (0.009)	0.005 (0.009)
Diaspora·Polity destin.	-0.145 (0.066)**	-0.143 (0.065)**	-0.147 (0.064)**
Diaspora·ln Distance	0.206 (0.044)***	0.206 (0.046)***	0.204 (0.046)***
Diaspora·colonial ties	-0.225 (0.115)*	-0.224 (0.115)*	-0.231 (0.114)**
70s decade dummy	14,626.044 (9,728.974)	12,196.643 (11,070.146)	6,250.142 (14,717.810)
80s decade dummy	20,531.777 (12,179.696)*	17,497.429 (14,157.532)	6,033.374 (18,997.611)
90s decade dummy	17,612.533 (12,555.358)	14,557.502 (14,583.544)	2,623.095 (19,812.158)
Hungary			-20,621.760 (5,024.035)***
Korea			10,987.328 (6,106.036)*
Libya			-11,259.023 (2,858.195)***
Philippines			10,243.047 (4,171.631)**
Poland			-17,519.094 (16,209.281)
Australia			21,235.941 (5,468.616)***
Austria			16,663.734 (5,582.218)***
Canada			12,312.768 (3,759.060)***
Denmark			21,662.011 (6,624.619)***
Finland			23,615.182 (6,958.391)***
Greece			15,026.878 (4,829.052)***
Ireland			27,816.129 (8,174.078)***
Netherlands			11,036.807 (4,042.974)***
New Zealand			36,203.913 (9,157.681)***
Norway			23,341.644 (7,102.956)***
Portugal			15,650.042 (5,564.595)***
Sweden			18,508.206

---

Switzerland			(5,593.966)*** 15,764.055 (6,252.543)**
Constant	-660,333.247 (481,855.158)	-469,627.004 (215,951.319)**	-803,757.282 (254,768.646)***
Observations	6379	6379	6379
R-squared	0.84	0.84	0.85

Notes: Dependent variable is migration flow, robust standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 4 presents different treatments for the standard errors for the core model plus country dummies. In the first column we cluster the standard errors by destination, in the second column by origin and in the third column by dyad. The significance levels of the variables are mostly unaffected by the different clustering strategies.

**Table 4: Core Model with Country Dummies**

	(1)	(2)	(3)
	SE clustered by destination	SE clustered by origin	SE clustered by dyad
Diaspora	5.192 (1.952)***	5.192 (1.552)***	5.192 (2.117)**
Colonial Ties	-517,111.006 (287,995.560)*	-517,111.006 (407,849.026)	-517,111.006 (299,053.554)*
ln Distance (km)	75,351.471 (23,091.526)***	75,351.471 (42,098.008)*	75,351.471 (24,076.083)***
Landlocked (dummy)	249.341 (666.920)	249.341 (480.775)	249.341 (467.698)
ln GDP origin	1,002.362 (561.803)*	1,002.362 (555.529)*	1,002.362 (495.685)**
ln GDP destination	69,372.786 (30,055.712)**	69,372.786 (37,819.368)*	69,372.786 (32,627.928)**
ln Population origin	441.954 (358.516)	441.954 (240.849)*	441.954 (264.451)*
ln Population destination	10,446.629 (3,203.039)***	10,446.629 (3,280.990)***	10,446.629 (3,334.947)***
Polity origin	308.563 (88.727)***	308.563 (143.616)**	308.563 (65.157)***
Polity destination	-145.203 (172.751)	-145.203 (249.328)	-145.203 (194.361)
Growth origin	466.350 (295.700)	466.350 (229.624)*	466.350 (280.171)*
Growth destination	-3,536.482 (1,855.674)*	-3,536.482 (2,580.077)	-3,536.482 (1,645.835)**
International War	2,841.200 (3,840.216)	2,841.200 (2,703.759)	2,841.200 (2,869.927)
Civil War	570.184 (1,161.129)	570.184 (1,304.884)	570.184 (942.841)
Relative GDP	12,364.304 (25,011.694)	12,364.304 (13,980.374)	12,364.304 (25,157.280)
70s decade dummy	6,250.142 (15,710.492)	6,250.142 (9,154.209)	6,250.142 (17,776.888)
80s decade dummy	6,033.374 (20,014.547)	6,033.374 (11,751.617)	6,033.374 (22,680.688)
90s decade dummy	2,623.095 (21,054.560)	2,623.095 (11,715.324)	2,623.095 (23,682.858)
Colonial·GDP origin	7,995.639 (4,280.311)*	7,995.639 (4,704.553)	7,995.639 (4,082.188)*
Colonial·GDP destin.	47,240.864 (27,866.710)*	47,240.864 (39,492.348)	47,240.864 (29,467.240)
Colonial·Growth origin	-4,648.869 (3,998.427)	-4,648.869 (3,648.589)	-4,648.869 (3,768.951)
Colonial·Growth destin.	26,606.456 (14,748.845)*	26,606.456 (19,783.430)	26,606.456 (14,816.239)*
Colonial·rel.GDP	-58,557.623 (33,272.297)*	-58,557.623 (45,488.957)	-58,557.623 (37,891.568)
Colonial·Polity origin	-616.944 (627.596)	-616.944 (372.219)	-616.944 (607.765)
ln Distance· lnGDP destin	-8,640.344 (2,508.189)***	-8,640.344 (4,603.891)*	-8,640.344 (2,585.768)***
ln Distance· Growth origin	0.047 (0.083)	0.047 (0.056)	0.047 (0.084)
Diaspora·relative GDP	0.829 (0.142)***	0.829 (0.068)***	0.829 (0.153)***
Diaspora·ln Distance	0.204 (0.030)***	0.204 (0.036)***	0.204 (0.043)***



Diaspora·colonial ties	-0.231 (0.103)**	-0.231 (0.064)***	-0.231 (0.108)**
Diaspora·Polity origin	0.005 (0.010)	0.005 (0.003)	0.005 (0.011)
Diaspora·Polity destin.	-0.147 (0.054)***	-0.147 (0.033)***	-0.147 (0.055)***
Diaspora·ln GDP origin	-0.189 (0.077)**	-0.189 (0.090)**	-0.189 (0.094)**
Diaspora·ln GDP destin.	-0.439 (0.166)***	-0.439 (0.129)***	-0.439 (0.170)**
Hungary	-20,621.760 (2,563.992)***	-20,621.760 (8,238.913)**	-20,621.760 (5,698.891)***
Korea	10,987.328 (3,539.862)***	10,987.328 (7,168.193)	10,987.328 (7,102.962)
Libya	-11,259.023 (2,698.750)***	-11,259.023 (4,521.476)**	-11,259.023 (3,201.855)***
Philippines	10,243.047 (2,231.414)***	10,243.047 (4,921.307)*	10,243.047 (4,909.327)**
Poland	-17,519.094 (6,637.319)***	-17,519.094 (15,548.581)	-17,519.094 (13,841.669)
Australia	21,235.941 (6,645.473)***	21,235.941 (9,095.708)**	21,235.941 (6,699.914)***
Austria	16,663.734 (5,865.056)***	16,663.734 (6,872.041)**	16,663.734 (6,407.450)***
Canada	12,312.768 (4,036.971)***	12,312.768 (5,574.620)**	12,312.768 (4,193.404)***
Denmark	21,662.011 (7,473.352)***	21,662.011 (8,490.405)**	21,662.011 (7,779.763)***
Finland	23,615.182 (8,060.104)***	23,615.182 (8,740.953)**	23,615.182 (8,350.251)***
Greece	15,026.878 (5,405.905)***	15,026.878 (5,407.324)**	15,026.878 (5,657.926)***
Ireland	27,816.129 (8,900.860)***	27,816.129 (8,881.375)***	27,816.129 (9,996.492)***
Netherlands	11,036.807 (4,068.661)***	11,036.807 (5,556.854)*	11,036.807 (4,565.343)**
New Zealand	36,203.913 (10,671.847)***	36,203.913 (12,597.448)***	36,203.913 (11,304.762)***
Norway	23,341.644 (7,785.718)***	23,341.644 (9,512.973)**	23,341.644 (8,243.978)***
Portugal	15,650.042 (4,781.613)***	15,650.042 (4,687.319)***	15,650.042 (6,875.743)**
Sweden	18,508.206 (6,314.486)***	18,508.206 (7,635.849)**	18,508.206 (6,557.772)***
Switzerland	15,764.055 (5,438.924)***	15,764.055 (8,501.205)*	15,764.055 (6,739.487)**
Constant	-803,757.282 (267,442.633)***	-803,757.282 (372,401.085)**	-803,757.282 (306,550.935)***
Observations	6379	6379	6379
R-squared	0.85	0.85	0.85

Notes: Dependent variable is migration flow, robust standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

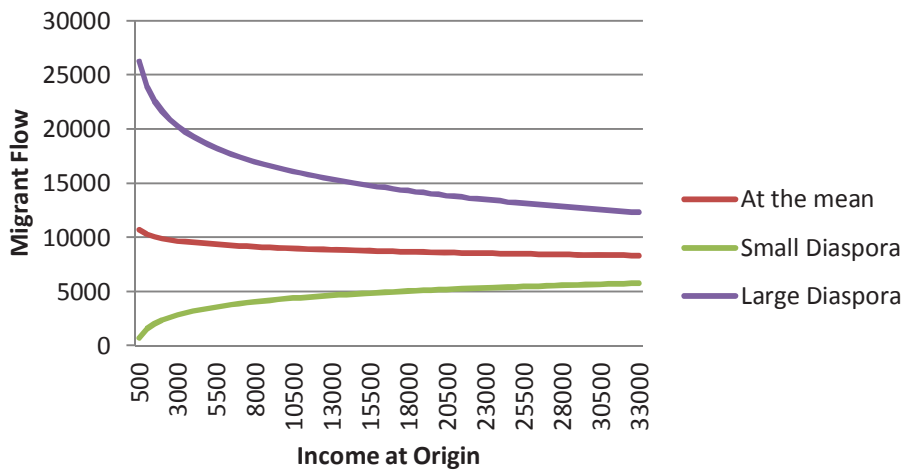
## Discussion

The overall influence of each variable is the sum of its direct and interaction effects. We find that interactions are central to the migration decision. Since it is difficult to discern from the regression tables the net effects of variables that operate through several interactions, we show the salient relationships through graphs.

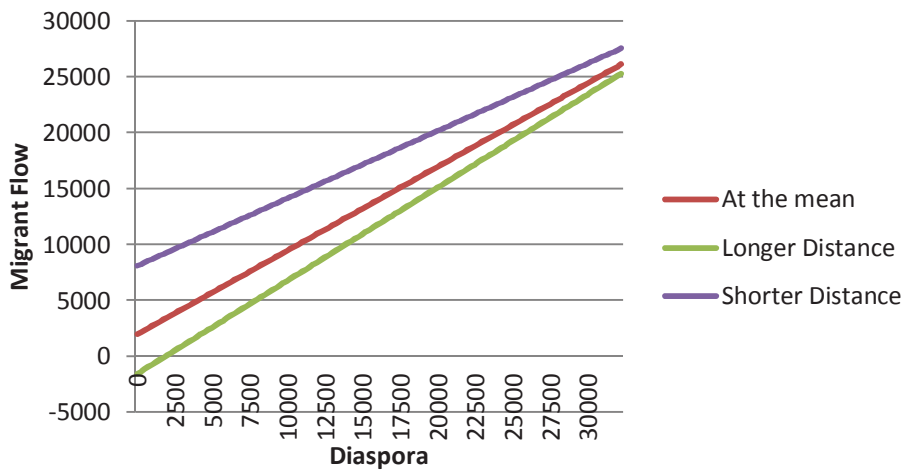
As suggested by Beine *et al* (2011a) the effect of the diaspora on migration flows is large. However, its overall effect is determined through a series of interactions with other variables. Of the 15 interaction terms included in our core model, eight are interactions with the diaspora. Figures 5-7 bring out the salient net effects. In Figure 5 the space is the flow of migration as a function of the level of income in the country-of-origin. The three lines show this relationship for different sizes of diasporas. The flat red line shows the relationship at the mean of all variables: income in the country-of-origin has little effect. This confirms the suggestions by De Haas (2007) that the level of development has little impact on the migration. The upward-sloping green line shows the relationship if the diaspora is small. Now, income in the country-of-origin is strongly supportive of migration: the ability to finance the investment cost evidently predominates over the incentive effect. Finally, the downward-sloping purple line shows the relationship if the diaspora is large. Now, higher income in the country-of-origin reduces migration: the incentive effect predominates over the enhanced ability to finance investment. Our interpretation of these substantial differences is that the diaspora is critical in reducing the investment cost that must be borne by the migrant, and so substitutes for income in the country-of-origin. A similar implication follows from Figure 6, which depicts the flow of migration as a function of the size of the diaspora. At the mean of all other variables (the red line) the diaspora substantially increases migration. The slope of this line is 0.749, suggesting that for an extra 1,000 persons in the diaspora, the flow in the following decade will increase by about 749 migrants. Although we use different data and employ a different model and estimation strategy, this diaspora effect is comparable to the one found by Beine *et al* (2011). They regress the change in the stock of migration between 1990 and 2000 on the stock of migrants in 1990. The coefficient on this lagged stock of migrants takes values of 0.669 to 0.766 (Their Table 1, column 3 and robustness checks in Table 2). This is a very large effect and an obvious concern is that since the diaspora is measured as the accumulated stock of past migration it spuriously includes all those unobserved effects which are time-invariant. This concern is mitigated because we are

controlling for fixed effects in both the country-of-origin and the host country; because we are controlling for the past relationship of the country dyad; because the effect of the diaspora is not direct but comes about only through interactions with other variables that have clear economic interpretations; and because they are consistent with the recent literature. The effect of the diaspora is considerably more powerful if migration requires a long distance (the green line) rather than a short distance (the purple line). Distance is an important impediment to migration, but since long distances increase the investment cost, diasporas have a more important role. Finally, Figure 7 depicts the same relationship but for former colonies (the blue line). A former colonial relationship reduces the importance of diasporas, our interpretation being that it substitutes for the information flow provided by diasporas.

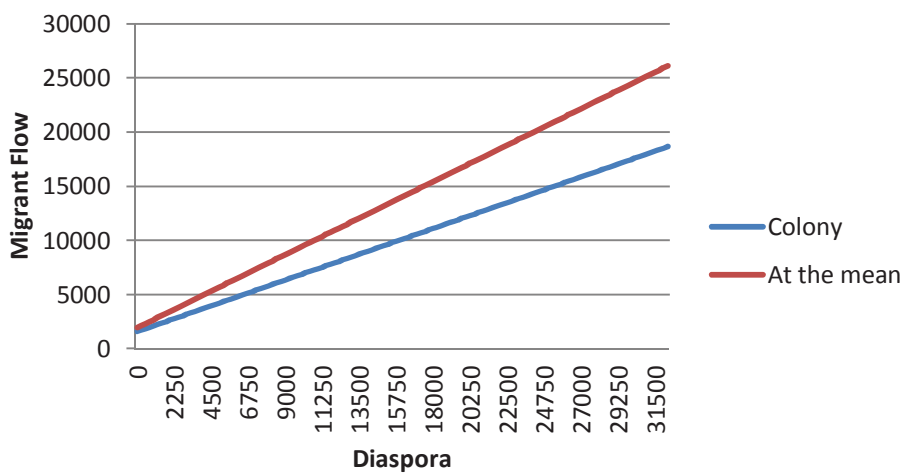
**Figure 5: Income, Diaspora and Migrant Flow**



**Figure 6: Diaspora and Migrant Flow**



**Figure 7: Colonial Past and Migrant Flow**



While a key attraction of OECD countries is evidently their level of income, the small variations between them and the small changes over time do not appear to be important. In our results, the significant and positive direct effect of the level of income is offset by interaction terms. Similarly, growth during the decade in the destination country has only one weakly significant effect, through an interaction with a former colonial relationship. Growth during the decade in the country-of-origin is also weakly significant, increasing migration. This is consistent with the effect on the ability to finance investment predominating over the incentive effect. However, it should be recalled that we proxy the destination economy's growth rate by the weighted growth rates of its export partners.

To summarize the economic aspects of the migration decision, it appears to be determined by the tension between the prospect of reaping a large economic gain and the impediment of financing the required initial investment. Diasporas are critical in overcoming the investment costs and hence elevating the importance of the incentive constituted by the income gap.

Whereas the economic effects are powerful but complex, the effect of political openness is weak but straightforward. There are only small differences among OECD societies in measured political openness and these have small, though statistically significant effects. Political openness in countries-of-origin is more important, but greater openness increases emigration rather than reducing it. Evidently, variations in political rights include the right to emigrate and the right to information pertinent for the migration decision.

Turning to our proxies for variations in restrictions on migration, both a former colonial relationship and whether the country-of-origin is landlocked are significant. However, quantitatively the effects are minor, and that concerning whether the country-of-origin is landlocked is likely to be spurious since it appears to be driven entirely by Hungary.

### ***Treatment of the Standard Errors***

The columns of Tables 4 provide a comparison of the treatment of the standard errors: clustered by country of origin (1), country of destination (2) and by dyad (3)<sup>2</sup>. Comparison across the different standard error treatments suggests that our results are robust to different ways of clustering. In particular the coefficient on diaspora and the interaction terms remain unchanged. This is also the case for our core model, Appendix Table A3 provides the same comparison for the core model without the country dummies. There are some small differences as a result of different clustering methods when we compare columns (1) and (2). When we cluster the standard errors by origin (column 1) the colonial past dummy and two interaction terms are significant when compared with clustering by destination (2). Thus, once we allow for clustering of standard errors by destination, the colony effects become less important. In columns (2) and (3) the Korea and Poland dummies are insignificant, when we cluster by destination or by dyad these countries are not different from the average.

### **3. Migration and Cultural Distance**

We now introduce the possible effects of cultural difference into our analysis. In the economic literature Spolaore and Wacziarg (2009, 2013) have pioneered the use of the objective measurement of cultural distance between pairs of countries by means of the proxy of neutral genetic divergence.<sup>3</sup> They (2013:18) define genetic distance as a “summary measure of differences in allele frequencies between populations across a range of neutral genes (chromosomal loci) ... [it] captures the length of time since two populations became separated from each other. When two populations split apart, random genetic mutations result in genetic differentiation over time. The longer the separation time, the greater the genetic distance computed from a set of neutral genes. Therefore, genetic distance captures the time since two populations have shared common ancestors (the time since they were parts of the same population), and can be viewed as a summary measure of relatedness between populations. An intuitive analogue is the concept of relatedness between individuals: two siblings are more closely related than two cousins because they share more recent common

---

<sup>2</sup> We also considered correcting for dyadic correlation following Fafchamps and Gubert (2007) but in order to apply their method we would have to analyze dyads for the same countries of origin and destination. In our case we only consider migration from poor and middle income to rich countries and exclude migration from rich countries.

<sup>3</sup> The original genetic distance data were collected by Cavalli-Sforza, Menozzi, and Piazza (1994, pp. 75–76).

ancestors—their parents rather than their grandparents.” The measured genetic differences are neutral in the sense that they have no direct bearing upon human performance, but they are reasonable proxies for the time during which different pairs of human cultures have been separated and so have presumably diverged. Like Spolaore and Wacziarg (2009, 2013) we concentrate on the genetic distance that takes into account that a country’s population is not homogeneous but made up of different groups. We therefore use the “weighted genetic distance,” representing the *expected* genetic distance between two randomly chosen individuals, one from each country, using the genetic distances associated with their respective ancestor populations (Spolaore and Wacziarg, 2013:18 Footnote 17).

Spolaore and Wacziarg (2009) have used their measure of the cultural distance between countries to demonstrate that it can be a barrier to the flow of knowledge. While their concern is the adoption of new technologies, potentially the same knowledge barrier impedes the movement of people. We test this by introducing cultural distance as an explanatory variable. Our results are reported in Table 5. In column 1 we repeat our core model for ease of comparison. Our measure of the genetic distance between countries is not available for as many pairs of countries as our other data and so the number of observations is approximately halved. However, it remains substantial at 3,022 and when our core regression is rerun on this reduced sample the results are very similar: the size of the diaspora continues to have a large and significant effect. We report these results in column (2). At the mean of all the variables our core model predicts a flow of 9,760 migrants whereas the results from the reduced sample in column (2) predict one thousand more migrants (10,760). Having established the difference of predicted outcomes given different sample sizes we then introduce the genetic distance variable in column (3). We find no significant direct effect of genetic distance on the rate of migration. Migration from poor countries to rich ones appears not to be impeded by the gap in cultures.

However, cultural distance can potentially affect migration not only as a direct barrier, but indirectly as a barrier to the integration of immigrants into their host society. Potentially this can affect migration by reducing the pace at which the diaspora decays through absorption into mainstream society. This is, for example, an implication of the Dunbar constant, which posits that the number of other people with which an individual interacts is approximately the same across societies (Dunbar 1992). Recall that we have found further evidence for the proposition of Beine *et al* (2011a) that diasporas are critical adjuncts to the migration process.

As the diaspora accumulates, migration becomes easier and so increases. While migration adds to the diaspora, an offsetting process of absorption of immigrants into mainstream society reduces it as migrants gradually lose their connections with their country-of-origin.

This indirect effect of cultural distance on migration yields the counter-intuitive proposition that *the wider is cultural distance the more rapidly migration will accelerate*. Due to the lower rate of absorption, a given rate of migration will generate a more rapid accumulation of the diaspora, and this in turn will facilitate more migration. This hypothesis was proposed in Collier (2013) but was not tested.

We test the hypothesis through introducing an interaction term between the cultural distance between a host society and a country-of-origin and our previous estimate of the accumulated stock of migrants from that country. While we have referred to this cumulated stock as the diaspora, our measure has made no allowance for cultural absorption.

As reported in Table 5 column 3, whereas the direct effect of cultural distance on migration is small and insignificant, its interaction with the diaspora has a large and significant effect. As postulated, the wider is the gap in cultures the more useful is the accumulated stock of immigrants to subsequent migrants. The effect is substantial. At the mean value of genetic distance (a value of 1,040) the predicted flow of migrants during a decade is 12,066. If we consider a smaller genetic distance (reducing the distance by one standard deviation, 599) this predicted value is 8,347 and if we consider a larger genetic distance (plus one standard deviation) this predicted value is 15,784. As we noted in the previous section, while the apparent direct effect of the diaspora might be contaminated by the effect of omitted variables, interaction effects are less exposed to this critique. In the present case, we have the same measured diaspora having powerfully different effects on migration due to observed differences in cultural distance. If cultural distance affects migration by slowing the rate of decay of the diaspora, this is potentially discernible by decomposing accumulated migration into vintages. In Table 5 column 4 we use the information on migration flows to split the diaspora into the stock of migrants who have been living in their country of destination 20 years or longer ('old diaspora') and the more recently arrived ('new diaspora'). Consistent with our interpretation, the interaction of cultural distance with the diaspora is significantly larger for the older vintage of migrants than for the recent vintage.



**Table 5: Migration and Cultural Distance**

	(1)	(2)	(3)	(4)	(5)
	Core Model	Core Reduced Sample			
Diaspora	5.192 (2.146)**	8.072 (2.485)***	8.169 (3.259)**		11.824 (2.926)***
Colonial Ties	-517,111.006 (282,859.500)*	-622,813.483 (283,780.093)**	-445,214.096 (318,846.868)	-305,004.925 (237,477.007)	-608,305.259 (282,663.600)**
ln Distance (km)	75,351.471 (24,424.608)***	79,085.698 (22,128.864)***	50,014.560 (32,100.779)	-9,964.227 (23,751.349)	49,005.884 (20,784.402)**
Landlocked (dummy)	249.341 (488.219)	130.304 (609.999)	663.212 (864.367)	-233.551 (534.963)	455.115 (622.551)
ln GDP origin	1,002.362 (411.665)**	1,584.541 (450.606)***	2,687.055 (907.052)***	856.099 (491.705)*	1,515.574 (547.414)***
ln GDP destination	69,372.786 (27,979.365)**	48,604.599 (31,292.668)	36,865.992 (39,206.839)	9,551.756 (25,212.490)	20,376.770 (30,808.287)
ln Population orig.	441.954 (280.743)	632.072 (300.659)**	1,041.546 (415.213)**	407.116 (246.165)*	721.233 (280.732)**
ln Population dest.	10,446.629 (2,742.483)***	12,896.656 (2,474.915)***	12,502.308 (3,070.433)***	9,019.825 (3,220.997)***	10,818.214 (1,968.057)***
Polity origin	308.563 (70.860)***	384.530 (86.909)***	474.985 (117.067)***	230.175 (99.394)**	242.436 (79.236)***
Polity destination	-145.203 (174.725)	-233.171 (149.390)	-240.955 (186.623)	74.012 (175.246)	-56.087 (141.429)
Growth origin	466.350 (238.402)*	715.804 (305.135)**	630.072 (354.444)*	536.651 (278.002)*	692.279 (269.433)**
Growth destination	-3,536.482 (1,408.284)**	-6,773.273 (2,368.818)***	-7,680.048 (2,647.838)***	-4,584.929 (2,614.457)*	-4,909.115 (2,010.585)**
International War	2,841.200 (2,997.171)	-1,722.397 (1,602.142)	6,316.431 (5,550.783)	296.973 (2,119.653)	961.898 (1,707.000)
Civil War	570.184 (1,156.637)	1,294.382 (1,536.324)	1,608.248 (2,153.692)	2,543.530 (1,523.613)*	1,825.772 (1,560.941)
Relative GDP	12,364.304 (21,061.814)	38,648.389 (32,061.762)	23,591.108 (29,109.390)	-21,100.996 (15,568.825)	35,645.070 (30,152.097)
70s decade dummy	6,250.142 (14,717.810)	20,923.464 (22,217.917)	10,813.548 (21,173.244)	-16,404.083 (12,051.120)	21,499.741 (21,226.392)
80s decade dummy	6,033.374 (18,997.611)	24,272.016 (27,984.331)	9,195.668 (27,623.421)	-23,371.305 (15,041.416)	25,869.269 (26,751.396)
90s decade dummy	2,623.095 (19,812.158)	21,120.314 (29,306.061)	5,475.725 (28,885.053)	-24,673.526 (15,699.780)	23,903.574 (28,044.079)
Colonial-GDP orig	7,995.639 (3,646.708)**	7,173.645 (3,553.547)**	4,235.444 (3,768.915)	2,831.413 (3,487.534)	5,075.525 (3,698.789)
Colonial-GDP dest	47,240.864 (28,582.516)*	57,964.246 (28,448.333)**	43,836.482 (33,598.420)	29,208.996 (23,690.464)	56,876.785 (28,368.459)**
Colonial-Growth o	-4,648.869 (3,724.751)	-4,719.796 (3,849.393)	-6,126.260 (5,078.522)	-5,172.677 (2,736.166)*	-3,909.704 (3,563.158)
Colonial-Growth d	26,606.456 (14,310.639)*	28,640.508 (14,204.026)**	25,226.127 (18,102.862)	19,384.161 (11,781.072)*	28,338.936 (13,755.869)**
Colonial-rel.GDP	-58,557.623 (38,468.775)	-48,835.004 (38,421.453)	-61,595.417 (48,611.098)	-25,204.821 (34,581.298)	-41,106.514 (37,803.439)
Colonial-Polity o	-616.944 (538.178)	-526.889 (494.310)	39.392 (647.846)	721.512 (488.413)	-314.570 (497.287)
ln Dist. · lnGDP d	-8,640.344 (2,600.825)***	-9,103.128 (2,364.813)***	-6,237.354 (3,500.253)*	713.697 (2,498.729)	-5,795.870 (2,202.671)***
ln Dist · Growth o	0.047 (0.066)	-0.064 (0.065)	0.039 (0.091)	-0.030 (0.078)	-0.144 (0.067)**
Diaspora · rel.GDP	0.829 (0.133)***	0.847 (0.092)***	0.784 (0.113)***	0.531 (0.117)***	0.732 (0.086)***
Diaspora · ln Dist	0.204 (0.046)***	0.182 (0.072)**	0.047 (0.071)	-0.018 (0.049)	0.043 (0.095)
Diaspora · colonial	-0.231 (0.114)**	-0.425 (0.126)***	-0.301 (0.118)**	-0.238 (0.127)*	-0.497 (0.126)***
Diaspora · Polity o	0.005 (0.009)	0.009 (0.007)	0.004 (0.009)	-0.003 (0.008)	0.012 (0.008)
Diaspora · Polity d	-0.147 (0.064)**	-0.242 (0.080)***	-0.203 (0.059)***	-0.257 (0.093)***	-0.257 (0.081)***
Diaspora · ln GDP o	-0.189 (0.087)**	-0.307 (0.065)***	-0.278 (0.103)***	-0.246 (0.083)***	-0.332 (0.073)***
Diaspora · ln GDP d	-0.439 (0.197)**	-0.487 (0.208)**	-0.503 (0.269)*	0.145 (0.194)	-0.715 (0.244)***
Hungary	-20,621.760 (5,024.035)***	-22,048.182 (5,036.539)***	-27,456.013 (8,169.996)***	-8,374.042 (3,469.664)**	-21,365.605 (5,833.434)***
Korea	10,987.328 (6,106.036)*	12,562.352 (8,792.510)	12,939.121 (9,008.085)	11,758.153 (7,159.184)	11,927.027 (7,882.261)

Libya	-11,259.023 (2,858.195)***	-9,041.666 (2,279.608)***	-14,244.299 (5,171.100)***	-1,525.911 (1,443.534)	-6,670.643 (1,972.105)***
Philippines	10,243.047 (4,171.631)**	29,542.778 (12,038.574)**	17,853.394 (10,375.716)*	24,593.461 (11,890.278)**	27,050.676 (11,011.717)**
Poland	-17,519.094 (16,209.281)	-43,457.852 (19,009.933)**	-6,007.583 (26,661.787)	27,631.922 (23,497.565)	-18,506.735 (10,570.177)*
Australia	21,235.941 (5,468.616)***	29,441.281 (6,349.040)***	30,728.887 (7,448.069)***	20,365.119 (7,864.666)***	24,591.844 (5,076.969)***
Austria	16,663.734 (5,582.218)***	17,760.525 (4,653.381)***	15,612.821 (6,693.071)**	16,086.144 (6,649.154)**	15,544.078 (3,928.983)***
Canada	12,312.768 (3,759.060)***	16,105.362 (3,496.019)***	15,970.858 (4,120.182)***	11,838.892 (4,446.994)***	13,765.867 (2,916.846)***
Denmark	21,662.011 (6,624.619)***	27,595.584 (5,668.700)***	25,506.490 (7,412.917)***	20,668.605 (7,727.507)***	23,840.158 (4,751.281)***
Finland	23,615.182 (6,958.391)***	30,288.817 (6,169.385)***	27,626.736 (7,779.603)***	21,174.208 (7,951.920)***	25,804.660 (5,048.459)***
Greece	15,026.878 (4,829.052)***	17,301.370 (3,585.395)***	15,838.755 (5,205.319)***	15,390.434 (5,622.313)***	15,154.958 (3,011.156)***
Ireland	27,816.129 (8,174.078)***	33,234.418 (6,594.621)***	31,581.731 (8,902.163)***	25,384.600 (9,263.317)***	27,720.468 (5,339.430)***
Netherlands	11,036.807 (4,042.974)***	14,458.754 (3,249.670)***	11,554.071 (4,466.463)***	11,614.163 (4,548.343)**	12,785.090 (2,863.079)***
New Zealand	36,203.913 (9,157.681)***	45,854.514 (8,970.510)***	45,453.662 (10,900.044)***	31,370.859 (11,176.649)***	37,532.477 (7,057.826)***
Norway	23,341.644 (7,102.956)***	29,736.496 (5,961.339)***	27,572.539 (7,831.848)***	23,271.573 (8,101.697)***	25,872.279 (5,045.890)***
Portugal	15,650.042 (5,564.595)***	14,871.816 (4,972.856)***	14,104.181 (6,025.008)**	15,832.239 (5,517.851)***	10,209.207 (4,641.594)**
Sweden	18,508.206 (5,593.966)***	24,099.706 (5,033.747)***	22,450.097 (6,255.110)***	17,836.133 (6,581.647)***	20,508.140 (4,113.414)***
Switzerland	15,764.055 (6,252.543)**	20,343.539 (4,596.857)***	19,851.809 (6,563.475)***	21,293.396 (6,804.803)***	17,331.975 (4,001.408)***
Genetic Distance			1.170 (1.090)	1.686 (0.966)*	0.414 (0.816)
Diaspora·Gen Dist			0.000 (0.000)**		0.001 (0.000)***
New Diaspora				18.570 (3.835)***	
New Dias·GenDist				-0.010 (0.003)***	
Old Diaspora				2.692 (2.637)	
Old Dias·GenDist				0.0004 (0.0002)**	
Linguistic Dist.					-937.481 (355.333)***
Diaspora·LingDist.					0.003 (0.019)
Constant	-803,757.282 (254,768.646)***	-675,890.043 (268,210.198)**	-533,520.758 (345,671.431)	-199,193.396 (233,078.783)	-375,955.486 (263,162.534)
Observations	6379	3022	3394	3274	3022
R-squared	0.85	0.96	0.88	0.95	0.96

Notes: Dependent variable is migration flow, robust standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

While the above table show the single-decade effect of cultural distance, the impact of the diaspora on migration cumulates over several decades. In our core model, adding 100 migrants to the stock in 1960 would cumulate by 2000 to 823 immigrants as the higher initial stock increases the flow in the next decade, which in turn increases the stock, in a process that echoes down the decades. Re-estimated on the reduced sample for which it is feasible to include genetic distance this baseline effect is somewhat larger at 1247. Around this baseline if we repeat the comparison of plus or minus one standard deviation variation in genetic distance the cumulated stocks of migrants by 2000 are respectively 2213 and 637. Thus,

setting all other variables to the same values, a culturally distant country-of-origin will over the course of 40 years have sent approximately triple the number of immigrants to a host country as a culturally proximate one.

We now turn to robustness checks. We use an alternative measure of genetic difference and control for linguistic distance which we present in the appendix<sup>4</sup>. The results remain qualitatively similar. In column 5 we include a measure of linguistic distance. These data were also used by Spolaore and Wacziarg (2009) but were originally compiled by Fearon (2003). Using ethnographic data from Ethnologue, Fearon classifies languages into common families and describes the relatedness of world languages in a language tree.<sup>5</sup> Languages develop over time and space. If two languages share many common nodes in a historic language tree, these languages are more likely to trace their roots to a more recent common ancestor language. The number of common nodes in the linguistic tree is the measure of linguistic similarity. We converted this measure of similarity into a measure of genetic distance. We include this linguistic distance measure and its interaction with the diaspora (column 5). The correlation between genetic and linguistic distance is relatively low at  $\rho=0.24$ , thus the two variables measure different things.<sup>6</sup>

Linguistic difference has quite different effects from cultural distance, mirroring some of the recent results by Adserà and Pytliková (2012). Whereas a wide culture gap does not directly impede migration, a wide language gap is a significant direct impediment. There are several potential routes by which this may operate: for example, information flows may be more limited, or anticipated earning power in a host society may be reduced. However, the width of the language gap does not significantly interact with the diaspora. Again this is in contrast to the effect of the culture gap. Perhaps once in the host country, the rate at which the migrants forge connections with their host society and loosen them with their country-of-origin is determined by cultural distance rather than by language proximity. Once two languages are mutually unintelligible, the absorption of the new language may follow from the culturally-

---

<sup>4</sup> Cavalli-Sforza, Menozzi, and Piazza (1994) offer alternative measurements of genetic distance, mainly through weighting population groups differently.

<sup>5</sup> See <https://www.ethnologue.com>, accessed 31<sup>st</sup> March 2014.

<sup>6</sup> This is also confirmed in the recent work by Alesina et al (2013) where they examine the differential effect of birthplace, ethnic (linguistic) and genetic diversity on output.

determined frequency of social interaction with the indigenous population rather than being set by its intrinsic difficulty.

The contrast between the effects of cultural and linguistic difference is striking. Based on the regression presented in column 5, a one standard deviation increase in cultural distance *increases* the migration flow by 27 per cent, whereas a one standard deviation increase in linguistic distance *reduces* the migration flow by 16 per cent. Clearly, these distinctive effects warrant further research, including complementary studies at the micro-level.

#### **4. Conclusion**

Recent advances in global data sets have greatly enhanced the scope for the quantitative analysis of international migration. We have focused on migration from developing to developed countries. During the first half of the 20<sup>th</sup> Century such migration was negligible, but since 1945 it has continuously increased and is now substantial. By using global data since 1960 we are able to analyse this process of escalation. Developed and developing countries are characterized by three substantial differences: in incomes, in political rights, and in cultures. There are now objective measures for each of these differences and our analysis is able to integrate them. We find that migration is driven by interactions between a few key variables.

One powerful effect is that diasporas increase migration. Most plausibly this is through reducing the investment costs that migrants must surmount in order to access the markedly higher incomes that can be earned in OECD countries. For example, diasporas are more important if migration requires a long journey, and if incomes are low in the country-of-origin. As migration feeds diasporas and diasporas facilitate subsequent migration, this accounts for the observed tendency of migration to accelerate.

The other powerful effect is from cultural differences. Counter-intuitively, a wide gap between the culture of the host society and that of the country-of-origin increases rather than impedes migration. The migration-enhancing effect is solely due to an interaction between cultural distance and the accumulated stock of migrants: the wider the gap the more powerfully does past migration facilitate further migration. We have interpreted this as being due to the effect of cultural difference on the rate at which migrants are absorbed into their

host society and hence loosen their connections with their country-of-origin. This interpretation is amenable to validation through sociological studies. More generally, now that the scope for global aggregate analysis has opened up, there is a need for close integration with field studies.

## Appendix

Adjustment of Migrant Stocks for Mortality to calculate adjusted Migrant Flows:

For the US we have information on the different age groups of the migrant stock for the years 1960, 1970, 1980 and 1990. We know the total number of migrants and how many migrants were under one years of age, between the ages of 1-4, 5-9, ..., 80-84 and how many were over 85 year olds. We only have this detailed information for other OECD migrant stocks for 1990 and 2000.

The UN population statistics provide information on the mortality rates by age group for (almost all) OECD countries for the years 1960, 1970, 1980 and 1990. The data set list the probability of dying for infants under one years of age, and the probability of death between the ages of 1-4, 5-9, ..., 80-84 and for the over 85 year olds. The probabilities are for death occurring during the following ten years.

For each OECD country we use their mortality rates by age group and the information on the US migrant stock and Since we constructed the adjustments by using US migrant stock information, it is useful to consider how much smaller the migrant stock at time  $t$ ,  $M_{it}$ , would have been due to these deaths. These are the percentage reductions for the US and for the average OECD country:

### A1 Adjustment of Migrant Stock Figures

	1960-69	1970-79	1980-89	1990-99
Average	10.74	10.36	6.99	4.50
US	10.71	9.85	6.73	4.61

We used these percentages to adjust the stock at time  $t$ .

### A2 Background Information: US migrant stock by age group (in per cent)

	1990	1980	1970	1960
under 15	7.51	8.84	6.31	5.18
15-44	56.83	48.35	34.70	24.70
45-64	22.02	21.64	27.03	37.48
over 65	13.64	21.17	31.97	32.64

**Table A3: Core Model Without Country Dummies – Standard Errors Clustered**

	(1)	(2)	(3)
	SE clustered by destination	SE clustered by origin	SE clustered by dyad
Diaspora	4.911 (1.976)**	4.911 (1.544)***	4.911 (2.112)**
Colonial Ties	-524,780.577 (293,711.053)*	-524,780.577 (417,120.639)	-524,780.577 (307,420.660)*
ln Distance (km)	55,277.536 (20,409.709)***	55,277.536 (35,341.383)	55,277.536 (22,030.398)**
Landlocked (dummy)	-1,428.186 (1,201.211)	-1,428.186 (809.246)*	-1,428.186 (570.351)**
ln GDP origin	252.857 (595.129)	252.857 (247.192)	252.857 (479.913)
ln GDP destination	46,405.190 (24,429.034)*	46,405.190 (31,585.665)	46,405.190 (25,739.420)*
ln Population origin	413.798 (371.257)	413.798 (256.592)	413.798 (302.399)
ln Population destination	1,968.244 (562.299)***	1,968.244 (828.569)**	1,968.244 (520.525)***
Polity origin	280.284 (82.242)***	280.284 (128.488)**	280.284 (63.308)***
Polity destination	228.528 (117.657)*	228.528 (265.003)	228.528 (122.259)*
Growth origin	422.381 (257.878)	422.381 (186.419)**	422.381 (252.567)*
Growth destination	1,259.936 (662.069)*	1,259.936 (1,165.049)	1,259.936 (748.105)*
International War	2,288.745 (4,004.183)	2,288.745 (2,714.017)	2,288.745 (3,008.148)
Civil War	1,204.098 (1,262.386)	1,204.098 (1,561.947)	1,204.098 (1,024.703)
Relative GDP	11,270.338 (19,334.723)	11,270.338 (13,043.731)	11,270.338 (18,963.625)
70s decade dummy	12,196.643 (12,513.150)	12,196.643 (9,070.876)	12,196.643 (13,021.468)
80s decade dummy	17,497.429 (16,231.897)	17,497.429 (12,081.918)	17,497.429 (16,424.064)
90s decade dummy	14,557.502 (16,876.163)	14,557.502 (11,106.185)	14,557.502 (16,886.483)
Colonial·GDP origin	8,275.607 (4,326.667)*	8,275.607 (4,800.454)	8,275.607 (4,293.613)*
Colonial·GDP destin.	47,239.411 (28,368.738)*	47,239.411 (40,244.663)	47,239.411 (30,128.278)
Colonial·Growth origin	-4,838.823 (4,134.177)	-4,838.823 (3,792.251)	-4,838.823 (3,915.004)
Colonial·Growth destin.	27,272.263 (15,145.438)*	27,272.263 (20,519.647)	27,272.263 (15,398.477)*
Colonial·rel.GDP	-56,274.398 (33,936.937)	-56,274.398 (47,259.337)	-56,274.398 (38,722.356)
Colonial·Polity origin	-778.772 (626.388)	-778.772 (440.886)*	-778.772 (607.777)
ln Distance· lnGDP destin	-6,246.882 (2,196.248)***	-6,246.882 (3,787.064)	-6,246.882 (2,347.377)***
ln Distance· Growth origin	0.056 (0.081)	0.056 (0.052)	0.056 (0.082)
Diaspora·relative GDP	0.819 (0.139)***	0.819 (0.067)***	0.819 (0.148)***
Diaspora·ln Distance	0.206 (0.032)***	0.206 (0.037)***	0.206 (0.045)***

Diaspora·Col. ties	-0.224 (0.104)**	-0.224 (0.071)***	-0.224 (0.110)**
Diaspora·Polity o	0.005 (0.010)	0.005 (0.003)*	0.005 (0.011)
Diaspora·Polity d	-0.143 (0.055)**	-0.143 (0.033)***	-0.143 (0.057)**
Diaspora·lnGDP o	-0.190 (0.077)**	-0.190 (0.091)*	-0.190 (0.094)**
Diaspora·lnGDP d	-0.415 (0.170)**	-0.415 (0.129)***	-0.415 (0.172)**
Constant	-469,627.004 (217,460.958)**	-469,627.004 (301,280.426)	-469,627.004 (232,195.431)**
Observations	6379	6379	6379
R-squared	0.84	0.84	0.84

Notes: Dependent variable is migration flow, robust standard errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%



**Table 5A: Migration and Cultural Distance – Alternative Measure of Cultural Distance**

	(1)	(2)	(3)	(4)
	Core Model			
Diaspora	5.192 (2.146)**	8.800 (3.342)***		12.452 (2.826)***
Colonial Ties	-517,111.006 (282,859.500)*	-450,145.851 (317,027.699)	-289,555.678 (244,958.016)	-627,372.769 (281,267.116)**
ln Distance (km)	75,351.471 (24,424.608)***	48,731.609 (32,102.759)	-14,460.844 (24,402.310)	45,271.857 (20,321.745)**
Landlocked (dummy)	249.341 (488.219)	678.666 (868.064)	-230.450 (563.608)	489.920 (631.978)
ln GDP origin	1,002.362 (411.665)**	2,555.080 (862.298)***	826.631 (467.900)*	1,468.444 (513.977)***
ln GDP destination	69,372.786 (27,979.365)**	35,913.827 (39,523.437)	8,654.440 (25,915.472)	17,955.262 (30,505.067)
ln Population orig.	441.954 (280.743)	1,003.886 (418.782)**	467.660 (261.257)*	705.450 (275.156)**
ln Population dest.	10,446.629 (2,742.483)***	11,902.555 (3,057.558)***	8,935.470 (3,260.408)***	10,681.726 (1,956.140)***
Polity origin	308.563 (70.860)***	448.435 (112.571)***	217.354 (96.805)**	228.738 (76.318)***
Polity destination	-145.203 (174.725)	-215.079 (184.491)	85.010 (171.914)	-54.397 (141.290)
Growth origin	466.350 (238.402)*	605.042 (345.033)*	516.030 (277.316)*	674.495 (260.915)***
Growth destination	-3,536.482 (1,408.284)**	-7,363.863 (2,594.587)***	-4,527.261 (2,660.533)*	-4,840.696 (1,963.124)**
International War	2,841.200 (2,997.171)	6,355.797 (5,447.929)	911.428 (2,274.388)	1,060.331 (1,689.182)
Civil War	570.184 (1,156.637)	1,564.822 (2,165.913)	2,477.633 (1,522.138)	1,792.709 (1,577.388)
Relative GDP	12,364.304 (21,061.814)	22,642.839 (29,029.084)	-24,520.137 (16,104.415)	34,100.497 (29,255.131)
70s decade dummy	6,250.142 (14,717.810)	10,556.395 (21,088.413)	-18,935.217 (12,420.506)	20,439.883 (20,670.974)
80s decade dummy	6,033.374 (18,997.611)	8,949.119 (27,499.312)	-26,812.590 (15,589.220)*	24,471.750 (26,072.114)
90s decade dummy	2,623.095 (19,812.158)	5,382.375 (28,755.173)	-28,251.663 (16,257.359)*	22,512.678 (27,358.650)
Colonial·GDP orig	7,995.639 (3,646.708)**	4,967.943 (4,002.111)	2,922.762 (3,699.521)	6,634.704 (3,865.805)*
Colonial·GDP dest	47,240.864 (28,582.516)*	43,532.491 (33,376.576)	27,308.903 (24,468.137)	57,268.238 (28,056.533)**
Colonial·Growth o	-4,648.869 (3,724.751)	-5,597.152 (5,032.455)	-5,195.973 (2,828.438)*	-3,518.767 (3,474.940)
Colonial·Growth d	26,606.456 (14,310.639)*	24,743.592 (17,962.158)	19,318.004 (12,155.916)	28,549.071 (13,567.109)**
Colonial·rel.GDP	-58,557.623 (38,468.775)	-61,387.961 (48,476.297)	-22,921.285 (35,355.106)	-40,482.939 (37,108.987)
Colonial·Polity o	-616.944 (538.178)	1.276 (673.719)	724.332 (490.330)	-286.996 (489.376)
ln Dist· lnGDP d	-8,640.344 (2,600.825)***	-6,057.278 (3,506.240)*	1,214.585 (2,566.893)	-5,366.073 (2,150.339)**
ln Dist· Growth o	0.047 (0.066)	0.026 (0.092)	-0.037 (0.082)	-0.155 (0.063)**
Diaspora·rel.GDP	0.829 (0.133)***	0.799 (0.109)***	0.523 (0.115)***	0.761 (0.081)***
Diaspora·ln Dist	0.204 (0.046)***	0.027 (0.070)	-0.050 (0.052)	0.006 (0.092)
Diaspora·colonial	-0.231 (0.114)**	-0.314 (0.117)***	-0.244 (0.123)**	-0.511 (0.125)***
Diaspora·Polity o	0.005 (0.009)	0.003 (0.009)	-0.002 (0.008)	0.011 (0.008)
Diaspora·Polity d	-0.147 (0.064)**	-0.219 (0.055)***	-0.270 (0.090)***	-0.258 (0.079)***
Diaspora·ln GDP o	-0.189 (0.087)**	-0.295 (0.104)***	-0.257 (0.087)***	-0.349 (0.074)***
Diaspora·ln GDP d	-0.439 (0.197)**	-0.522 (0.272)*	0.133 (0.202)	-0.747 (0.234)***
Hungary	-20,621.760 (5,024.035)***	-23,893.285 (7,539.779)***	-5,699.006 (2,996.577)*	-18,230.961 (4,609.313)***

Korea	10,987.328 (6,106.036)*	11,678.080 (8,624.139)	12,113.957 (7,482.153)	10,861.805 (7,550.656)
Libya	-11,259.023 (2,858.195)***	-13,709.264 (5,159.047)***	-1,560.187 (1,469.746)	-6,105.132 (1,907.763)***
Philippines	10,243.047 (4,171.631)**	16,566.216 (9,961.582)*	23,036.320 (11,429.554)**	26,372.196 (10,779.761)**
Poland	-17,519.094 (16,209.281)	-4,909.066 (27,540.724)	32,042.066 (24,454.259)	-15,522.723 (10,347.151)
Australia	21,235.941 (5,468.616)***	29,081.006 (7,301.241)***	19,453.018 (7,767.452)**	24,115.276 (4,932.315)***
Austria	16,663.734 (5,582.218)***	14,395.410 (6,609.967)**	15,366.561 (6,567.510)**	15,451.282 (3,824.057)***
Canada	12,312.768 (3,759.060)***	15,107.519 (4,045.991)***	11,327.803 (4,402.896)**	13,539.588 (2,851.535)***
Denmark	21,662.011 (6,624.619)***	24,066.995 (7,356.511)***	20,077.366 (7,716.524)***	23,574.498 (4,696.353)***
Finland	23,615.182 (6,958.391)***	26,694.920 (7,830.444)***	21,494.141 (8,255.280)***	25,736.114 (5,032.382)***
Greece	15,026.878 (4,829.052)***	14,993.461 (5,203.251)***	15,413.958 (5,728.925)***	15,131.619 (3,002.475)***
Ireland	27,816.129 (8,174.078)***	29,742.102 (8,833.368)***	25,050.778 (9,282.156)***	27,398.053 (5,335.252)***
Netherlands	11,036.807 (4,042.974)***	10,834.157 (4,430.338)**	11,094.211 (4,527.189)**	12,748.836 (2,838.807)***
New Zealand	36,203.913 (9,157.681)***	43,157.805 (10,821.574)***	30,697.570 (11,214.261)***	36,872.449 (7,004.436)***
Norway	23,341.644 (7,102.956)***	26,039.912 (7,767.389)***	22,604.047 (8,093.191)***	25,562.923 (4,997.922)***
Portugal	15,650.042 (5,564.595)***	12,931.725 (5,970.457)**	15,858.161 (5,553.380)***	9,963.527 (4,661.787)**
Sweden	18,508.206 (5,593.966)***	21,218.075 (6,203.771)***	17,272.048 (6,580.877)***	20,259.836 (4,060.258)***
Switzerland	15,764.055 (6,252.543)**	18,683.031 (6,440.570)***	20,876.174 (6,769.026)***	17,238.158 (4,000.381)***
Genetic Distance		4.358 (5.331)	6.743 (4.772)	1.753 (4.028)
Diaspora·Gen Dist		0.003 (0.001)***		0.003 (0.001)***
Linguistic Dist.				-981.403 (356.786)***
Diaspora·LingDist.				0.011 (0.019)
New Diaspora			17.739 (3.856)***	
New Dias·GenDist			-0.053 (0.020)***	
Old Diaspora			3.270 (2.792)	
Old Dias·GenDist			0.003 (0.001)**	
Constant	-803,757.282 (254,768.646)***	-515,522.368 (347,557.909)	-186,840.917 (238,406.287)	-350,363.410 (261,638.857)
Observations	6379	3394	3274	3022
R-squared	0.85	0.89	0.95	0.96

Note: The alternative measure of cultural distance is based on giving different weights to the genetic groups.

## Descriptive Statistics

Variable	Mean	Std. Dev.	Minimum	Maximum
Migration Flow	6245.989	81429.41	-767413.1	4920606
Diaspora	10399.08	101620.3	0	4662233
Colonial Ties	.0487537	.2153696	0	1
Distance (km)	7921.693	3787.909	238	19109
Landlocked	.1831008	.3867795	0	1
GDP origin	4155.399	5319.776	339.7033	46139.15
GDP destination	17820.77	6232.604	4071.565	33074
Population origin	3.48e+07	1.24e+08	163000.1	1.15e+09
Population destintation	3.57e+07	5.18e+07	2371747	2.50e+08
Polity origin	-1.612589	6.286122	-10	10
Polity destination	9.063333	3.213823	-9	10
Growth origin	3.07756	1.501785	-1.044421	9.748272
Growth destination	2.188784	.9161661	1.025059	5.575115
International War	.0583164	.234359	0	1
Civil War	.1498668	.3569687	0	1
Relative GDP	.857292	.251403	.3770696	2.205505

#### Origin Countries:

Afghanistan, Albania, Algeria, Angola, Argentina, Bahrain, Bangladesh, Benin, Bolivia, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Cape Verde, Central African Republic, Chad, Chile, China, Columbia, Comoros, Democratic Republic of the Congo, Republic of Congo, Cote d'Ivoire, Cuba, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Fiji, Gabon, Gambia, Ghana, Guatemala, Guinea, Guinea Bissau, Guyana, Haiti, Honduras, Hungary, India, Indonesia, Iran, Iraq, Jamaica, Jordan, Kenya, Korea, Laos, Liberia, Libya, Madagascar, Malawi, Malaysia, Mali, Mauretania, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Romania, Rwanda, Saudi Arabia, Senegal, Sierra Leone, Solomon Islands, Somalia, Sri Lanka, Sudan, Syria, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, Uruguay, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

#### Destination Countries:

Australia, Austria, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States.

## References

- Adserà, A. and P. Mariola. 2012. The Role of Language in Shaping International Migration. IZA Discussion Paper No. 6333.
- Alesina, A., Harnoss, J. and H. Rapoport. 2013. Birthplace Diversity and Economic Prosperity. IZA Discussion Paper, No. 7568.
- Beine, M., F. Docquier and Ç. Özden, (2011a). Diasporas. *Journal of Economic Development* 95(1): 30-41.
- Beine, M., F. Docquier and Ç. Özden. (2011b). Diaspora Effects in International Migration. Key Questions and Methodological Issues. World Bank Policy Research Working Paper 5721.
- Beine, M., S. Bertoli and J Fernandez-Hueras Moraga. 2014. A practitioner's guide to gravity models of international migration. Mimeo.
- Bertoli, S and J Fernandez-Hueras Moraga. 2013. Multilateral resistance to migration. *Journal of Development Economics* 102: 79–100.
- Brettell, C.B. and Hollifield, J.F. (eds). 2008. *Migration Theory: Talking Across Disciplines*, 2nd edition. New York: Routledge.
- Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes* (Princeton, NJ: Princeton University Press).
- Collier, P. 2013. *Exodus: Immigration and Multiculturalism in the 21st Century*. Oxford University Press: Oxford.
- Dunbar, R.I.M. 1992. Neocortex size as a constraint on group size in primates, *Journal of Human Evolution* 22: 469-493.
- De Haas, H. 2010. *Migration Transitions: A Theoretical and Empirical Enquiry into the Developmental Drivers of International Migration*. International Migration Institute, University of Oxford, IMI Working Paper 24/2010.
- De Haas, H. 2007. Turning the tide? Why development will not stop migration. *Development and Change* 38(5): 819-840.
- Doornik, J. 2009. Autometrics in Castle, J.L. and N. Shephard (eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Oxford University Press, Oxford, UK.
- Docquier, F. and Rapoport, H. 2012. Globalisation, Brain Drain, and Development, *Journal of Economic Literature* 50 (3): 681-730.
- Fearon, J. 2003. Ethnic and Cultural Diversity By Country. *Journal of Economic Growth* 8:195–222.

- Fafchamps, M. and Gubert, F. 2007. The formation of risk sharing networks. *Journal of Development Economics* 83(2): 326–350.
- Hatton, T.J. and Williamson, J.G. 1998. *The Age of Mass Migration: Causes and Economic Impact*. New York: Oxford University Press.
- Johansen, S. and Nielsen, B. 2009. An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator in Castle, J.L. and N. Shephard (eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Oxford University Press, Oxford, UK.
- Koopmans, R., I. Michalowski and S. Waibel. 2012. Citizenship Rights for Immigrants: National Political Processes and Cross-National Convergence in Western Europe, 1980-2008. *American Journal of Sociology* 117(4):1202-1245.
- Lee, E.S. 1966. A theory of migration. *Demography* 3: 47–57.
- Özden, Ç., Parsons, C. R., Schiff, M. and T.L. Walmsley. 2011. Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960–2000, World Bank Economic Review 25(1): 12-56.
- Ravenstein, E.G. 1885. The laws of migration. *Journal of the Royal Statistical Society* 48: 167–227.
- Ravenstein, E.G. 1889. The laws of migration. *Journal of the Royal Statistical Society* 52: 214–301.
- Spolaore, E. and R. Wacziarg. 2009. The Diffusion of Development. *Quarterly Journal of Economics* 124 (2): 469–529.
- Spolaore, E. and R. Wacziarg. 2013. How Deep are the Roots of Human Development? *Journal of Economic Literature* 51: 1–45.