

Dissecting Network Externalities in International Migration*

Michel Beine^a, Frédéric Docquier^b and Caglar Ozden^c

^a University of Luxembourg, Luxembourg, and CES-Ifo, Germany

^b FNRS and IRES, Université Catholique de Louvain, Belgium

^c World Bank, Development Research Group, United States

This preliminary and incomplete draft - June 2010

Abstract

Existing migrant networks play an important role in explaining the size and structure of immigration flows. They affect the net benefits of migration for future migrants by lowering assimilation costs (self-selection) and increase the probability of potential migrants to obtain a visa through family reunification programs (policy-selection). This paper presents an identification strategy allowing to disentangle these two channels. Then, it provides an empirical illustration based on US immigration data by metropolitan area and country of origin. First, we show that network externality is strong (elasticity of migration flows to network size is around one) and only a quarter of this elasticity is accounted for by the policy channel. Second, the policy-selection channel was stronger in the nineties than in the eighties due to more generous family reunion program. Third, the global elasticity and its policy contribution are greater for low skilled migrants.

*We thank Cristina Ileana Neagu for remarkable assistance. The second author acknowledges financial support from the ARC convention on "Geographical Mobility of Factors" (convention 09/14-019). The findings, conclusions and views expressed are entirely those of the authors and should not be attributed to the World Bank, its executive directors or the countries they represent.

1 Introduction

Existing diasporas or networks are among the most important determinants of the size, skill composition and destination choices of new migrants. The main channels through which diaspora effects influence migration patterns may be divided into two general categories. The first one is the overcoming of legal barriers and they help the migrant at the border *before* he arrives at the final destination. Diaspora members who already acquired citizenship or certain residency rights in the destination countries become eligible to sponsor their immediate families and other relatives. These family reunification programs are the main routes for many potential migrants in most OECD countries. The second channel is the lowering of assimilation costs which generally matter *after* the migrant crosses the border and settles at a specific geographic area in the destination country. Assimilation costs cover a wide range of obstacles faced by the migrants in finding employment, deciphering foreign social and cultural norms and adjusting to a new linguistic and economic environment. The informational, social and financial support provided by the existing network can be crucial.¹

The overall effect of diasporas have been clearly recognized in the sociology, demography and economics literatures and extensively analyzed over the last twenty years (such as Boyd, 1989). In terms of the assimilation costs, Massey et al. (1993) was one of the earliest papers, showing show diasporas reduce moving costs, both at the community level (e.g. inflow of people from the same nation helps creating subcultures), and family level (increase utility of friends and relatives). As shown by Carrington, Detragiache and Viswanath (1996), this explains why the size and structure of migration flows gradually change over time. In addition, networks provide information and assistance to new migrants before they left and when they arrive which facilitates newcomers' integration in the destination economy or reduces uncertainty. Based on a sample of individuals originating from multiple communities in Mexico and residing in the U.S., Munshi (2003) showed that an individual is more likely to be employed and earn higher wage when her network is larger.² Most recently,

¹Bauer et al. (2007) or Epstein (2008) argued that network effects might reflect her behavior in the sense that migrants with imperfect information about foreign locations follow the flow of other migrants, based on the (wrong or right) supposition that they had better information.

²On the contrary, Piacentini (2010) used data on migration and education from a rural region of Thailand to show that networks negatively affect the propensity of young migrants to pursue schooling while in the city.

exploiting a bilateral data set on international migration by educational attainment from 195 countries to 30 OECD countries in 1990 and 2000, Beine et. al. (2010) explore how diasporas affect the size and human capital structure of future migration flows. They find that the diasporas are by far the most important determinants, explaining over 70 percent of the observed variability of the size of flows. Regarding selection, diasporas were found to benefit the the migration of low-skilled relative to the highly-skilled, thus exerting a a negative impact and explaining over 45 percent of the variability of the selection ratio. using micro-data from Mexico, McKenzie and Rapoport (2007) frin the same effect which is also supported by Winters et.al (2001).

In terms of the effect of diasporas in overcoming policy induced migration restrictions, family reunification is the main legal route for many potential migrants in most continental European countries. Even in one of the most selective country such as Canada, about 40 percent of immigrants come under the family reunification and refugee programs, rather than selective employment or skill-based programs. Jasso and Rosenzweig (1986, 1989) estimated that each U.S. labor-certified immigrant generated a first-round multiplier around 1.2 within ten years (i.e. sponsored 0.2 relatives). Using a longer perspective, Bin Yu (2007) showed that each newcomer generates an additional inflow of 1.1 immigrant.

The goal of this paper is to empirically decompose the relative importance of these two channels - lowering of assimilation costs and overcoming policy induced legal barriers. The obvious natural approach is to directly use micro data on the various entry paths migrants use as well as their individual characteristics. Appropriate use of indicators on migration policies along with diaspora characteristics could provide information on the relative importance of kinship based admission of new migrants. Unfortunately, there is, to the best of our knowledge, no large micro database providing detailed information on the various entry tracks migrants use as well as the corresponding flows for each track. Furthermore, information on changes in immigration laws might not be enough to gauge the importance of family reunification policies over time. A good example is provided by regularization programs such as those that took place in the US in the nineties. As illegal migrants became legal residents, they became eligible to bring their close relatives to the US. This resulted in an increase in the number of migrants coming through family reunification in spite of no significant change in US migration laws.

As an alternative to the use of data on individual immigration paths, this paper

develops a different identification strategy using the migration data available at the city level for the United States.³ As mentioned earlier, the role of the diasporas in overcoming migration barriers operates at the border before the migrant settles in a given city. Thus, the probability for a migrant to obtain a visa through a family reunification program depends on the *total size* of the network already living in the United States, not on the distribution of the diaspora across different cities. On the other hand, the assimilation effect is mostly local and matters after the migrant chooses a city. For example, if a migrant lives in New York, the diaspora in Los Angeles is less likely to be of much help to him in terms of finding a job, especially relative to the network present in New York itself. This is the distinction we exploit to identify the relative importance of these two channels. We develop a simple theoretical model showing that, under plausible functional homogeneity of the two network externalities, the two different channels can be identified using U.S. census data on immigration by country of origin and by metropolitan area of destination. We then provide several extensions based on educational differences, time dimension, alternative migrant definitions or geographic areas and control of potential endogeneity.

We first show that the overall network effect is strong; the elasticity of migration flows to networks is around one, a result in line with Bin Yu (2007) and Beine et.al. (2010). Second, only a quarter of this elasticity is accounted for by the policy channel; the rest is due to the assimilation effect. Each immigrant sponsors 0.25-0.30 relative within ten years, a result in line with Jasso and Rosenzweig (1986, 1989). This shows the difficulty for host country government to curb the dynamics of immigration and confine multiplier effects. Third, the policy-selection channel was higher in the nineties than in the eighties due to more generous family reunion programs. Fourth, the global elasticity and its policy contribution are greater for low skilled migrants. Finally, these results are robust to the specification, to the choice of the dependent variable, to the definition in the relevant network and to the instrumentation of network sizes.

The remainder of this paper is organized as following. Section 2 uses a simple labor migration model to explain out identification strategy. Data are described and

³The US Census data is actually disaggregated at the metropolitan area level which might include multiple cities or a city and its surrounding areas. For simplicity, we use the phrase "city" instead of "metropolitan area."

econometric issues are discussed in Section 3. Results are provided in Section 4. Finally, Section 5 concludes.

2 Identification strategy

We use a simple model of labor migration where individuals with heterogeneous skill types s ($s = 1, \dots, S$) born in origin country i ($i = 1, \dots, I$) decide whether to stay in i or emigrate to location j ($j = 1, \dots, J$) in the destination country, the United States. The set of destination locations J share the same immigration policy but differ in other individual attributes. The individual utility is linear in income but also depends on possible migration and assimilation costs and characteristics of the city of residence. The utility of a type- s individual born in country i and staying in country i is given by:

$$u_{ii}^s = w_i^s + A_i^s + \varepsilon_{ii}$$

where w_i^s denotes the expected labor income in location i , A_i denotes country i 's characteristics (amenities, public expenditures, transfers, climate, etc.) and ε is a iid extreme-value distributed random term. The utility obtained when the same person migrates to location j is given by

$$u_{ij}^s = w_j^s + A_j^s - C_{ij}^s - V_{ij}^s + \varepsilon_{ij}$$

As in Beine et al. (2010), two types of migration costs are distinguished. On the one hand, C_{ij} captures moving and assimilation costs that are borne by the migrant. Together with $(w_j^s + A_j^s) - (w_i^s + A_i^s)$, C_{ij}^s would determine the net benefit of migration in a world of perfect labor mobility and the self-selection of migrants. We will assume below that C_{ij} depends on the network size in that specific location j . On the other hand, V_{ij}^s represents policy induced costs borne by the migrant to overcome the legal hurdles set by the destination country's government's (policy-selection channel). If a family reunion program is implemented at destination, V_{ij}^s also depends on the network size at the country level, not at the city level. Obviously, the main motivation to differentiate between these two types of costs is to identify the role of immigration policy on the size and structure of migration flows.

For simplification, we slightly abuse the terminology and refer to C_{ij}^s as *moving/assimilation costs* and to V_{ij}^s as *visa costs*. It is worth noticing that we allow

these costs to vary with skill type. It is well documented that high-skill workers are better informed than the low skilled, have higher capacity to assimilate or have more adaptive skills and, thus, face lower moving costs. In addition, the skill type also affects visa costs if there are selective immigration programs (such as points-system in Canada, Australia, New-Zealand, UK, the H1-B program in the US, etc.).

Let N_i^s denote the size of the native population of skill s that is within migration age in country i . When the random term follows an iid extreme-value distribution, we can apply the results in McFadden (1974) to write the probability that a type- s individual born in country i will move to location j as

$$\Pr \left[u_{ij}^s = \max_k u_{ik}^s \right] = \frac{N_{ij}^s}{N_i^s} = \frac{\exp [w_j^s + A_j^s - C_{ij}^s - V_{ij}^s]}{\sum_k \exp [w_k^s + A_k^s - C_{ik}^s - V_{ik}^s]},$$

and the bilateral "migrant-stayer" ratio from country i is given by

$$\frac{N_{ij}^s}{N_{ii}^s} = \frac{\exp [w_j^s + A_j^s - C_{ij}^s - V_{ij}^s]}{\exp [w_i^s + A_i^s]}$$

Hence, the log ratio of emigrants in city j to residents (N_{ij}^s/N_{ii}^s) is given by the following expression

$$\ln \left[\frac{N_{ij}^s}{N_{ii}^s} \right] = (w_j^s - w_i^s) + (A_j^s - A_i^s) - (C_{ij}^s + V_{ij}^s) \quad (1)$$

Let us now formalize network externalities. As stated above, both C_{ij}^s and V_{ij}^s depend on the existing network size. Moving/assimilation costs depends on origin country and host location characteristics (denoted by c_i^s and c_j^s respectively), increases with bilateral distance between i and j , and decreases with the size of the network at destination, M_{ij} (captured by the number of people living in location j and born in country i). In line with other empirical studies, we assume logarithmic form for distance and diaspora externality, and add one to the network size to get finite moving costs to destination where the network is null. This leads to

$$C_{ij}^s = c_i^s + c_j^s + \delta^s \ln d_{ij} - \alpha^s \ln (1 + M_{ij}) \quad (2)$$

where all parameters ($c_i^s, c_j^s, \delta^s, \alpha^s$) are allowed to vary with skill type s .

Regarding visa costs, we stated earlier that all cities share the same migra-

tion/border policy. An individual migrant's probability to cross the border using the family reunification program depends on the aggregate size of the network in these locations,

$$M_i \equiv \sum_{j \in J_1} M_{ij}.$$

Assuming the same logarithmic functional form for the network externality, the visa cost to each particular location included in J can be written as

$$V_{ij}^s = v_i^s - \beta^s \ln(1 + M_i) \quad (3)$$

where v_i^s stands for origin country characteristics, and the network externality β^s is allowed to vary with skill type.

Inserting (2) and (3) into (1) gives

$$\ln N_{ij}^s = \mu_i^s + \mu_j^s - \delta^s \ln d_{ij} + \alpha^s \ln(1 + M_{ij}) + \beta^s \ln(1 + M_i) \quad (4)$$

where $\mu_i^s \equiv \ln N_{ii}^s - w_i^s - A_i^s - c_i^s - v_i^s$ and $\mu_j^s \equiv w_j^s + A_j^s - c_j^s$ are combination of origin country' and destination location's characteristics,⁴ and (α^s, β^s) are the relative contributions of self-selection and policy-selection channels to the network externality.

It is obvious that regressing (4) using data on migration flows from the set of origin countries I to the set of locations J sharing common immigration policies cannot be used to identify the magnitude of the policy channel since $\ln(1 + M_i)$ is common to all destinations. It will simply be absorbed by the country fixed effects. However, we can take advantage of the identical functional form of the self-selection and policy-selection externalities to solve this problem. Focusing on the set of destinations J , the aggregate stock can be rewritten as $M_i = M_{ij} + \sum_{k \neq j} M_{ik}$. It follows that $\ln(1 + M_i)$ in (4) can be expressed as

$$\ln(1 + M_i) \equiv \ln(1 + M_{ij}) + \ln(1 + \Pi_{ij})$$

where $\Pi_{ij} \equiv (1 + M_{ij})^{-1} \sum_{k \neq j} M_{ik}$. Since we have migration and diaspora data available for the full set of locations in J , Π_{ij} can be constructed for each ij pair.

Assuming both externalities are linear (as in Pedersen et al., 2008, or McKenzie

⁴In principle, N_{ii}^s should be treated as an endogenous variable. We disregard this problem by assuming that each bilateral migration flow N_{ij}^s has a small impact on N_{ii}^s .

and Rapoport, 2007) or follow an homogenous function of degree a (e.g. M^a) enables us to perform this transformation. As a result, we can rewrite (4) as

$$\ln N_{ij}^s = \mu_i^s + \mu_j^s - \delta^s \ln d_{ij} + (\alpha^s + \beta^s) \ln (1 + M_{ij}) + \beta^s \ln (1 + \Pi_{ij}) \quad (5)$$

where β^s can be properly identified since Π_{ij} is a real bilateral variable. μ_i^s and μ_j^s capture all origin country and destination specific fixed effects. We can only properly estimate coefficients $\alpha^s + \beta^s$ and β^s from the above equation. However, the self-selection (assimilation) mechanism α^s might be recovered from subtracting β^s from $\alpha^s + \beta^s$.

3 Data

The data in this paper comes from 5% samples of the US Censuses of 1980, 1990 and 2000, which include detailed information on the social and economic status of foreign-born people in the United States. Of this array of information, we utilize characteristics such as gender, education level, country of birth and location in the US identified by metropolitan area. For the diaspora variable, we use all migrants in a given metropolitan area as reported in the 1990 census (or the 1980 census in the relevant sections). For the migration flow variable, we use the number of migrants (depending on the relevant definition) who arrived between 1990 and 2000 according to the 2000 census (or who arrived during 1980-1990 according to 1990 census). We re-group the educational variable provided by the US Census (up to 15 categories in the 2000 Census) to account for only 3 categories: up to Grade 11 (including no education), high-school graduate level (Grade 12), and some college or more. Since an indicator of the location where education was obtained is not available, we infer one distinguishing between the US versus home-country acquired education based on the age at which the immigrant reports to have entered the US. More specifically, we designate individuals as “US educated” if they arrived in the US before they would have normally finished their declared education level. For example, if a university graduate arrived at the age of 23 or older, then he/she is considered “home educated.”

The spherical distances used in this paper were calculated using STATA software based on geographical coordinates (latitudes and longitudes) found on the web: www.mapsofworld.com/utilities/world-latitude-longitude.htm, for country capital cities and www.realestate3d.com/gps/latlong.htm as well as Wikipedia for US

cities.7

3.1 Econometric issues

Equation (5), supplemented by an error term ϵ_{ij}^s , forms the basis of the estimation of the network effects. The structure of the error term can be decomposed in a simple fashion as:

$$\epsilon_{ij}^s = \nu_{ij}^s + u_{ij}^s \quad (6)$$

where u_{ij}^s are independently distributed random variables with zero mean and finite variance. And ν_{ij}^s reflect unobservable factors affecting the migration flows.

There are a couple of estimation issues raised by the nature of the data and the specification. These issues lead to inconsistency of usual estimates such as OLS estimates. An important issue is related to the high prevalence rate of zero values for the dependent variable N_{ij}^s . Depending on the period (1980's or 1990's), this proportion is between 50 and 70 percent of the total observations. Such large number of zero observations is fully consistent with our model which might arise due to the joint combination of small origin countries and small metropolitan areas between which distances and other costs would be prohibitive.

The high proportion of zero values has also been observed in trade data and creates similar problems of estimation. The use of the log specification drops the zero observations which constraints the estimation to a subsample involving only the country-city pairs for which we observe positive flows. This in turn tends to underestimate the key parameters α^s and β^s . One usual solution to that problem is take $\ln(1+N_{ij}^s)$ as the dependent variable and to estimate (5) by OLS. This makes the use of the global sample possible. Nevertheless, this adjustment is subject to a second statistical issue, i.e. the correlation of the error term u_{ij}^s with the covariates of (5). Santos-Silva and Tenreyro (2006) specifically cover this problem and propose some appropriate technique that minimizes the estimation bias of the parameters. This issue has also been addressed by Beine et al.(2009) in the context of international migration flows.

Santos-silva and Tenreyro (2006) show indeed that if the variance of u_{ij}^s depends on c_j^s , m_i^s , d_{ij} or M_{ij} , then its expected value will also depend on some of the regressors, invalidating one important assumption of consistency of OLS estimates. Furthermore, they show that the inconsistency of parameter estimates is also found

using alternative techniques such as (threshold) Tobit or non linear estimates. In contrast, in case of heteroskedasticity and a significant proportion of zero values, the Poisson Pseudo Maximum Likelihood (hereafter Poisson) estimates generates unbiased estimators of the parameters of (5). Furthermore, the Poisson estimates is found to perform quite well under various heteroskedasticity patterns and under rounding errors for the dependent variable. Therefore, in the subsequent estimates of (5), we use the Poisson estimation techniques and report the estimates for α^s , β^s and δ^s .

4 Results

We first estimate (5) with Poisson Pseudo Maximum Likelihood function. We use origin country and destination city fixed effects to capture the variables μ_i^s and μ_j^s respectively. We initially ignore skill differences by performing the estimation with aggregate migration flows. Then, we let coefficients vary by education level (sub-section 4.2) and account for education quality at origin (sub-section 4.3). Finally, we present a large set of robustness checks.

4.1 Global network externalities

In our first estimation, we assume that the coefficients $(\mu_i^s, \mu_j^s, \delta^s, \alpha^s, \beta^s)$ are homogenous across skill groups. The dependent variable N_{ij} in (5) measures the total migration flows from country i to U.S. metropolitan area j between 1990 and 2000. As explained above, the Poisson estimator takes care of the issues raised by the presence of large number of zeros for the migration flows. Nevertheless, the use of the full sample involves the inclusion of very small countries with idiosyncratic migration patterns. Many of these countries have fewer than a total of 500 migrants in the United States which results in a disproportionate amount of zero values for country-city pairs. Following Card (2008), we adjust the initial sample and leave out small countries which we define in terms of the total size of their diaspora in the US. We use different threshold values of this criterion : 10000, 7300, 5000, 2900 and 1040 US migrants. Those threshold correspond to key percentages in terms of included countries : 45, 50, 55, 60 and 70 percent of included countries. The results are reported in Table 1.

The global estimate of the diaspora effect is in line with previous results, for

instance with those of Beine et al. (2009). The key parameters are quite stable across subsamples which is partly due to the fact that, although we leave out a significant number of origin countries, the subsample capture a very high proportion of the US migrants - 98.8 percent in our smallest sample. We find that a one percent increase in the initial stock of diaspora leads to approximately one percent increase in the bilateral migration flow over a period between 1990 and 2000, given by the coefficient of $\alpha + \beta$. The results suggest that the diaspora effect is made up of about one fourth by the policy effect ($\frac{\beta}{\alpha+\beta}$) and three fourth by the pure network effect ($\frac{\alpha}{\alpha+\beta}$). In addition, the effect of the distance is also quite consistent with a coefficient of around -0.5, regardless of the sample size. In the subsequent estimations, we exclude countries with less than 7300 migrants.

Table 1. Overall Network Effects - per sub-samples

Parameters	Minimal size of total US diaspora				
	1040	2900	5000	7300	10000
$\alpha + \beta$	0.964	0.964	0.964	0.965	0.965
β	0.259	0.254	0.253	0.250	0.247
δ	-0.510	-0.498	-0.496	-0.490	-0.483
# observations	32807	27346	26862	25168	23958
% included countries	70	60	55	50	45
% included US migrants	99.9	99.7	99.5	99.2	98.8
Country FE	yes	yes	yes	yes	yes
City FE	yes	yes	yes	yes	yes

Notes: ML Poisson estimates of equation (1). All parameters significant at the 1 percent level; otherwise mentioned; robust estimates; Estimation carried out on migrants aged 15 and over, on the 1990-2000 period; threshold in terms of the size of the total diaspora at destination (across all U.S. metropolitan areas).

4.2 Education level

We now estimate (5) by education level since a large set of studies suggest diasporas are strong selection devices for migrants, but the effect declines with education. Microeconomic data (Mc Kenzie and Rapoport, 2009) show that the diaspora effect is smaller with the reported education level of the migrants. Beine et al. (2009) confirm the same result using bilateral data at the country level. In line with those studies,

we distinguish between reported education levels but we only focus the education that was obtained in the origin countries. In other words, we only include migrants who completed their education prior to migration and did not receive any further education in the United States. Obviously, this involves that there are potentially a lot of variability in the quality of education across origin countries. In order to account for such differences, we will take into account the level of development of the origin countries as a proxy of education quality in the next sub-section.

The migration data from the US census allow to account for the education level of the migrants, though with certain level of imperfection. The dependent variable is the number of migrants aged over 15 at time of their arrival and who completed their education at home. In the raw data, the education level reported by the migrants is given in number of (completed) schooling years. As usual in the literature, we classify education levels into 3 different categories : (i) low skilled migrants are those with less than 11 schooling years; (ii) medium skilled are those with more than 11 schooling years up to some high school degree; (iii) the high skilled migrants are those with some college degree or more. For some origin developing countries, the number of skilled migrants is quite limited. We estimate (5) for these three education levels separately and the results presented in Table 2.

In line with previous results, we find that the total diaspora effect ($\alpha+\beta$) decreases with the education level of migrants, from 1.146 for low skilled to 0.884 for high skilled migrants. Comparing skilled and unskilled migrants, we find the assimilation/network effect is higher for low skilled migrants relative to skilled migrants 0.763 vs. 0.655. The difference in the policy effect is, however, much more significant - 0.383 vs. 0.229. These results indicate that the diasporas are much important for the low skilled migrants in overcoming legal barriers in both relative and absolute terms.

Table 2. Results - Education level

Parameters	Education levels		
	Low skilled	Medium skilled	High skilled
$\alpha + \beta$	1.146	0.905	0.884
β	0.383	0.149	0.229
δ	-0.778	-0.452	-0.493
# obs	25168	25168	25168
Country FE	yes	yes	yes
City FE	yes	yes	yes

Notes: ML Poisson estimates of (5) on countries with less than 7300 migrants
All parameters significant at the 1 percent level, otherwise mentioned; robust estimates.

4.3 Education quality

As mentioned before, the results above include migrants who have obtained their education at home. It might therefore be desirable to account for the quality of education of the migrants in the estimation. It is very likely that the human capital quality of some migrants (especially if they are from poorer countries) can be quite low, even though their education levels are high on paper. If the perceived quality of education is low, the ability to migrate outside family reunification programs or other legal channels might be low. In that case, one could expect the visa effect and the network effects to be higher.

There is no common measure of quality of education by origin country. Nevertheless, using some measures of the observed skills for immigrants in Canada that obtained their education at home, Coulombe and Tremblay (2008) are able to estimate some skill-schooling gap. This measure provides some measure of the quality of education relative to the one prevailing in Canada. They also show that this measure is highly correlated with the level of GDP per head in the origin country. In line with this approach, we condition the diaspora effect on the World Bank classification dummies. In particular, we estimate (5) following the WB classification in terms of income while continuing to use some threshold in terms of minimal size of the US diaspora. These groups are: (i) low income countries, (ii) middle income countries and high income countries.

Results are reported in Table 3. Once again, assuming that education quality increases with the income of the origin country, we find that the diaspora effect decreases with income level and education quality, from 1.905 for low income countries to 0.968 for high income countries. In line with the previous estimates of Table 2, we find that most of the variation is driven by the visa/policy effects. the effect of the diaspora size through the visa effect is for high-income countries is a miniscule 0.211. On the other hand, it is 0.439 for middle income and 1.173 for low income countries. These results show clearly that the diaspora exert an important role for migrants from low income countries in providing them legal access to the US. On the other hand, the assimilation effect shows almost no variation - it is 0.732 for low

income countries and 0.757 for high income countries. Finally, migrants are much more sensitive to distance as seen with the sharp decline in the coefficient of distance with income levels.

Table 3. Accounting for income level at origin

Parameters	All countries	Low income	Middle income	High income
$\alpha + \beta$	0.966	1.905	1.126	0.968
β	0.242	1.173	0.439	0.211
δ	-0.486	-1.364	-0.883	-0.171
Nobs	25168	2902	12826	10164
Country FE	yes	yes	yes	yes
City FE	yes	yes	yes	yes

Notes: ML Poisson estimates of (5) on countries with more than 7300 migrants

All parameters significant at the 1 percent level. If not mentioned, robust estimates.

Estimation carried out on migration flows of individuals aged 15 and over.

4.4 Distance thresholds

Given the importance of distance in migration and the significant variation we observe in the earlier estimations, it is desirable to make a distinction between origin countries in terms of their distance from the US. One of the reasons is that the patterns of migration from different origin countries depend on the distance and the accessibility to the US territory, especially to different cities, by different means of transportation. For instance, countries close to the US border, such as Mexico and Central American countries, will send more migrants to cities in the south compared to countries located outside the Americas.

In the subsequent regressions, we define far and close countries on the basis of the minimal distance to the US border. The cutoff is 6790 kilometers which is the median distance in terms of pairs of origin countries and US metropolitan areas. We consider three variants in terms of education of the migrants : the all education case , unskilled migrants and skilled migrants.

First, we find that distance play a much more important role for migrants coming from far away countries. the coefficient of the distance variable is significantly lower in absolute value when the origin countries are closer to the US and these tend to be Latin American and Caribbean countries. Second, the overall diaspora effect is

slightly higher when origin countries are far away but this is not statistically significant. However, there is a difference in terms of the composition. The visa effect is higher for further away countries while the assimilation effect is more important for closer countries.

We obtain more nuanced results when we look compare the importance of distance for different education levels. For unskilled migrants, distance seem to be a very significant deterrent to the extent that it becomes prohibitive. We find that for the unskilled migrants from distant countries, the policy effect is almost non-existing. On the other hand, for skilled migrants from similar countries, the visa effect is much stronger when compared to nearby countries. Finally, we see that the assimilation effect difference between distant and nearby countries disappear when we control for the skill level - the earlier difference is simply due to the skill composition of migrants. In other words, once the migrants pass the border and enter the US, the assimilation effect of the diaspora does not differ based on the country of origin.

Table 4. Close versus remote countries

Parameters	All skill types		Low skilled		High skilled	
	Close	Far	Close	Far	Close	Far
$\alpha + \beta$	0.970	1.060	1.152	0.952	0.890	1.115
β	0.218	0.368	0.336	0.067 ⁿ	0.231	0.513
δ	-0.331	-1.065	-0.648	0.308 ⁿ	-0.330	-1.490
Log likelihood						
Nobs	14762	10406	14278	9680	14278	9680
Country FE	yes	yes	yes	yes	yes	yes
City FE	yes	yes	yes	yes	yes	yes

Notes: ML Poisson estimates of (5) on the flow of migrants aged 15 and over from countries with more than 10,000 migrants in the US

All parameters significant at the 1 percent level, except those superscripted n (non significant). If not mentioned, robust estimates.

Cut off value to define far and close: 6790 kilometers.

4.5 Alternative dependent

All of the earlier estimation results of (5) were based on the flows of migrants aged over 15 at time of arrival. In this section, we use alternative definitions of migration flows

and show that our estimates are roughly similar across the various definitions. More specifically, we consider 6 alternative definitions. In the first column, the migrants are aged between 15 and 65 at time of arrival but who did not necessarily complete their education in their origin country. In the second column, we take migrants aged between 15 and 65 in 2000 and who arrived between 1990 and 2000. The last four columns include male migrants only. The third column reports the results of male migrants aged over 15 at at time of arrival and who have completed their education at home. The fourth and fifth column concern the men aged between 15 and 65 while the last column report the results for all males.

- *flows15a65* : people between 15 and 65 at time of arrival between 1990 and 2000
- *flows1565* : people between 15 and 65 between 1990 and 2000
- *males15ah* : males over 15 at time of arrival between 1990 and 2000 and who completed education at home
- *males1565* : males between 15 and 65 between 1990 and 2000
- *males15a65*: males between 15 and 65 at time of arrival between 1990 and 2000
- *all males*: males arriving between 1990 and 2000

The results suggest that our estimates are fairly robust to the choice of alternative measures of the migration flows. The main part of the variation across categories of migrants concern the policy effect (β coefficient). The assimilation effect (α coefficient) is fairly constant across the types of migrants. The policy effect is found to be slightly higher for men.

Table 5. Alternative definitions of migrants

Parameters	flows15a65	flows1565	males15ah	males15a65	males1565	all males
$\alpha + \beta$	0.876	0.863	0.978	0.912	0.892	0.861
β	0.190	0.168	0.296	0.263	0.232	0.174
δ	-0.488	-0.487	-0.510	-0.507	-0.510	-0.465
Nobs	23958	23958	23958	23958	23958	23958
Country FE	yes	yes	yes	yes	yes	yes
City FE	yes	yes	yes	yes	yes	yes

Notes: ML Poisson estimates of (5) on countries with more than 10,000 migrants in the U.S.

All parameters significant at the 1 percent level.

4.6 Flows in the 90's vs 80's

Our analysis in the previous sections has focused on the effect of the diasporas present in 1990 on the migration flows between 1990 and 2000. Our dataset include parallel measures for the migration patterns in the 1980's. It is useful to perform the same estimation on the flows observed in the 1980's to observe if there has been any important changes. One possibility is to combine observations from the 1980's with those from the 1990's and adopt a panel approach pooling the data from the two cross section. Nevertheless, it is very likely that the expected effects (α and β) will be different over time and prevent us from pooling our data.

While it is unclear for the assimilation effect (α), the US immigration policy has obviously changed between the 1980s and the 1990s. There are arguments that the family reunification effect has significantly become stronger between the 1980's and the 1990's. First and importantly, the US immigration act of 1990 clearly expanded opportunities for family reunification. Nevertheless, there are two additional aspects that are not directly related to the 1990 law but that exert important effect on the extent of family reunification. An important element is that immediate relatives of US citizens are not limited or capped under the law. Therefore, quotas for family reunification that are established in the law can be exceeded in practice if the number of immediate family members of US citizens is above that estimated in the law for a given year. With more migrants obtaining the US citizenship, there is a natural upward trend in the number of people coming under the family reunification scheme *sensu lato*. The second important feature is related to the amnesty or legalization programs undertaken in 1986 via the Immigration Reform and Control Act. As a large amount of undocumented migrants became legal migrants, they became eligible to bring their family through the legal channels. Those who became citizens afterwards were even able to bring their relatives through the uncapped channel. Therefore, if our estimates make sense, those policy developments suggest that the estimated β coefficient has increased between the 1980's and 1990's.

Table 6 reports the estimates obtained for the 1990's and the 1980's. For each

period, we consider three alternative types of migrants : those with low education level, those with high education and all migrants. Our estimates suggest that the family reunification effects are uniformly stronger for the 1990's than for the 1980's for all types of migrants. Naturally, the change is more important for unskilled migrants, more than doubling within a decade. This is in line with the effects associated to legalization programs which concern primarily with undocumented migrants. In short, the comparison between the 1980's and the 1990's shows that our estimation of the policy effect is in line with what is expected from the evolution of the US immigration policy. On the other hand, the assimilation effect α stays around 0.75 for low skilled and 0.65 for high-skilled migrants across both decades.

Table 6. Flows in the 90's vs 80's

Parameters	90's			80's		
	All	LS	HS	All	LS	HS
$\alpha + \beta$	0.965	1.146	0.884	0.829	0.935	0.768
β	0.247	0.383	0.229	0.083	0.199	0.137
δ	-0.483	-0.778	-0.493	-0.580	-0.971	-0.527
Nobs	23958	25168	25168	20230	20300	20300
Country FE	yes	yes	yes	yes	yes	yes
City FE	yes	yes	yes	yes	yes	yes

Notes: All = All skill types; LS = low skilled; HS = high skilled

ML Poisson estimates of (5) on countries with more than 7300 migrants

All parameters significant at the 1 percent level ; robust estimates;

Estimation carried out on migration flows of individuals aged 15 and over.

4.7 The reflection problem

An important additional econometric issue is generated by the presence of unobserved bilateral factors ν_{ij}^s influencing the bilateral migration flows N_{ij}^s . In absence of observations for those factors, their effect will be included in the composite error term given by $\nu_{ij}^s + u_{ij}^s = \eta_{ij}^s$.⁵ If those factors also influence the diaspora M_{ij} , this

⁵Note that the non observation of ν_{ij}^s is also due to the fact that our data is of cross sectional nature. In fact, if one could introduce the time dimension in (5), one could estimate ν_{ij}^s through bilateral fixed effects. In our case, the use of time through a panel data framework is not possible because of the clear rejection of the pooling assumption. In fact, it is obvious that some parameters such as the one capturing the visa effect (β^s) are not constant over time. In the robustness analysis,

leads to some correlation between the error term and one covariate, invalidating the use of OLS (and Poisson) estimators. This case is likely to prevail in the context of this study, at least for a subset of observations. For instance, it might be the case that climate variables such as average temperature or average rain falls will affect the choice of migrants coming from some countries. This is likely to be the case for contemporaneous migrants (i.e. the N_{ij}^s variable) and the previous ones (i.e. the M_{ij} variable). While it is difficult to assess whether the size of this effect is important or not, it is desirable to use some alternative estimation techniques that take care of that issue. As suggested by Tenreyro (2007), one possibility is to combine Poisson estimators with instrumental variables estimator. This estimation can be done in the GMM context. Dropping the s subscript for convenience of exposition and aggregating all explanatory variables c_j^s , m_i^s , d_{ij} and M_{ij} into the x_{ij} vector, the Poisson estimator γ solves the following moment condition:

$$\sum_{ij}^n [N_{ij} - \exp(x_{ij}\gamma)]x_{ij} = 0 \quad (7)$$

. In order to take care of the correlation between x_{ij} and η_{ij} , one can use the following GMM estimator denoted by ψ :

$$\sum_{ij}^n [N_{ij} - \exp(x_{ij}\psi)]z_{ij} = 0 \quad (8)$$

in which z_{ij} represent the vector of instruments, i.e. variables that are supposed to be correlated with M_{ij} but uncorrelated with N_{ij} and η_{ij}^s . In this robustness analysis, we rely on the GMM estimator ψ using two potential instruments. Those instruments are the variables $\ln(1 + M_{ij})$ and $\ln(1 + \frac{M_{ij}^-}{1+M_{ij}})$ observed in 1950, i.e. about 40 years before the observed diaspora in the benchmark regression. Those variables are well correlated with their values in 1990 (part of the stock of 1990 was already present in 1950). In contrast, the network and policy effects on the flows during the 1990's associated to migrants already present in 1950 are supposed to be quite limited. The main drawback of using such an instrument is that it leads to a change in the available sample. This is first due to the fact that the definition of origin countries

we document the change in the US migration policy and show that the β^s parameter changes between the 1980's and the 1990's.

and US metropolitan areas has significantly changed between 1950 and 1990. A second reason is due to the independence of many former colonies during the 50's and 60's.⁶ Therefore, in the robustness analysis, we show on comparable samples that our benchmark regressions relying on Poisson regressions are not affected by the potential correlation between M_{ij} and ν_{ij}^s . In other terms, we show that the estimates for γ and for ψ are quite close on identical samples.

In practice, we first reestimate the Poisson regressions and use those estimates as a benchmark with respect to the 'IV' (GMM) estimates. Table 7 report the estimates of the Poisson on the restricted sample (column 2) and of the combined Poisson and IV estimates *à la Tenreyro* in column 3 and 4. In column 3 we use one instrument only, i.e. $\ln(1 + \frac{M_{ij}^-}{1+M_{ij}})$ observed in 1950 while in column 4 we supplement the instrument set with $\ln(1 + \frac{M_{ij}^-}{1+M_{ij}})$ observed in 1950, too.⁷

The results show that our estimates are strikingly robust to the instrumentation procedure. Both the total diaspora effect and the estimated policy effect are very similar across estimation methods. They are also very similar regardless of the inclusion or not of $\ln(1 + \frac{M_{ij}^-}{1+M_{ij}})$ variable observed in 1950.

Table 7. Instrumenting network sizes

	Poisson	Poisson+IV	
	Poisson	IV	IV
$\alpha + \beta$	1.029	1.015	1.005
β	0.338	0.356	0.350
δ	-0.711	-0.750	-0.754
Nobs	23541	23541	23541
Country FE	yes	yes	yes
City FE	yes	yes	yes

Notes: First column : ML Poisson estimates of (5). Two last columns: GMM estimates.

All parameters significant at the 1 percent level ; robust estimates.

Estimation carried out on migration flows of individuals aged 15 and over.

Instrument for IV estimates in col 3 : local network size observed in 1950.

⁶For instance, all US migrants coming from former European colonies were identified as migrants coming from the colonizing country.

⁷Note that , we checked the robustness of the maximum likelihood estimator. Indeed, the use of the Pseudo Poisson Maximum Likelihood might lead to convergence problems and might generate spurious convergence. Following Santos Silva and Tenreyro (2010), the issue might be addressed through some iterative procedure dropping the insignificant fixed effects.

Instruments for IV estimates in col 4: local and national network sizes observed in 1950.

4.8 Geographic area

Our identification strategy rests on the definition of metropolitan areas of the US Census bureau which defines our local network/diaspora variable. As a result we have assumed that the local network that impacts a migrant’s assimilation migration costs resides within the same US metropolitan area as the migrant. In order to test the robustness of this particular decision, we change the definition of the geographic area corresponding to the local network. In particular, we consider that the M_{ij} variable is composed by the number of migrants from country i living in metropolitan area j as well as in neighboring metropolitan areas that are located within 100 miles from the center of j . In general, around 50 percent of the cases, this leads to an increase in the size of the network.

Table 8 provides the estimation results of this change in the geographic area definition. We find that both effects are roughly similar with the estimates of the benchmark regression, presented in the first column. The assimilation/network effect is almost the same (around 0.61) but the policy effect is somewhat lower than the benchmark regression. The share of the network effect is slightly higher than in the benchmark regressions.

Table 8. Alternative definition of local network

Parameters	metr. area	100 miles
$\alpha + \beta$	0.965	0.875
β	0.247	0.164
δ	-0.483	-0.442
Nobs	23958	23958
Country FE	yes	yes
City FE	yes	yes

Notes: ML Poisson estimates of (5).

All parameters significant at the 1 percent level; robust estimates.

Estimation carried out on migration flows of individuals aged 15 and over.

5 Conclusion

TBA

6 References

Bauer, Thomas, Gil Epstein, Ira N. Gang (2007). Herd Effects or Migration Networks? The Location Choice of Mexican Immigrants in the U.S. *Research in Labor Economics*, 26, 199-229.

Beine, Michel, Frédéric Docquier, Caglar Ozden (2010). Diasporas. *Journal of Development Economics*, forthcoming.

Bin, Yu (2007). *Chain Migration Explained: The Power of the Immigration Multiplier*. LBF Scholar Publishing: El Paso, U.S.

Boyd, Monica (1989). Family and Personal Networks in International Migration: Recent Developments and New Agendas. *International Migration Review*, 23 (3), 638-670.

de Meza, David (1987). The Migration Multiplier. *Bulletin of Economic Research*, 39 (3), 243-48.

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

Gang, Ira N. and Luis A. Rivera-Batiz (1994). Labor Market Effects of Immigration in the United States and Europe: Substitution vs. Complementarity. *Journal of Population Economics* 7, 157-175.

Jasso, Guillermina, Mark R. Rosenzweig (1986). Family reunion and the immigration multiplier: U.S. immigration law, origin-country conditions, and the reproduction of immigrants. *Demography*, 23 (3), 291-311.

Jasso, Guillermina, Mark R. Rosenzweig (1989). Sponsors, Sponsorship Rates and the Immigration Multiplier. *International Migration Review*, 23 (4), 856-888.

Manski, Charles F. (1993). Identification of Endogeneous Social Effects: the Relection Problem. *Review of Economic Studies*, 60 (3), 531-42.

Massey, Douglas S., Joaquim Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J. Edward Taylor (1993). Theories of international migration: Review and Appraisal." *Population and Development Review*, 19 (3), 431-466.

McFadden, Daniel L. (1984). *Econometric analysis of qualitative response models*.

In: Z. Griliches and M. Intriligator, eds., Handbook of Econometrics, Volume 2, Amsterdam. Elsevier/North-Holland.

McKenzie, David, Hillel Rapoport (2007). Self-selection patterns in Mexico-US migration: the role of migration networks. Review of Economics and Statistics, forthcoming.

Munshi, Kaivan (2003). "Networks in the modern economy: Mexican migrants in the US labor market. Quarterly Journal of Economics, 118 (2), 549-99.

Pedersen, Peder J., Mariola Pytlikova, Nina Smith (2008). Selection or Network Effects? Migration Flows into 27 OECD Countries, 1990-2000. European Economic Review 52 (7), 1160-1186.

Piacentini, Mario (2010). Enclaves and schooling choices of young migrants. Mimeo-graph: University of Geneva.

Santos Silva, J.M.C. and S. Tenreyro (2006), The Log of Gravity, Review of Economics and Statistics, 88 (4), 641-658.

Santos Silva, J.M.C and Tenreyro, S. (2010), On the Existence of the Maximum Likelihood Estimates for Poisson Regression, *Forthcoming* in Economic Letters.

Tenreyro, S. (2007), On the Trade Impact of Nominal Exchange Rate Volatility, Journal of Development Economics, 82, 2, 485-508.

Winters, Paul, Alain de Janvry, Elisabeth Sadoulet (2001). Family and Community Networks in Mexico-U.S. Migration. Journal of Human Resources, 36 (1), 159-184