FACULTY OF LAW, ECONOMICS AND FINANCE

CREA
Discussion
Paper
2010-14

Center for Research in Economic Analysis
University of Luxembourg

Diaspora effects in international migration: key questions and methodological issues

available online : http://wwwfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers/2010

Michel Beine, University of Luxembourg

Frédéric Docquier, Université Catholique de Louvain

Caglar Ozden, DECRG, The World Bank

November 2, 2010

For editorial correspondence, please contact : elisa.ferreira@uni.lu

University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faïencerie
L-1511 Luxembourg

Diaspora effects in international migration: key questions and methodological issues.

Michel Beine[a], Frédéric Docquier[b] and Caglar Ozden[c]
[a]University of Luxemburg and CES-Ifo
[b]FNRS and IRES, Université Catholique de Louvain
[b]DECRG, The World Bank

November 2 2010

Abstract

This paper reviews the existing literature on the impact of migrants networks on the patterns of international migration. It covers the theoretical channels at stake in the global effect of the networks. It identifies the key issues, namely the impact on size, selection and concentration of the migration flows. The paper also reviews the empirical hurdles that the researchers face in assessing the importance of networks. The key issues concern the choice of micro vs a macro approach, the definition of a network, the access to suitable data and the adoption of econometric methods accounting for the main features of those data. Finally, the paper reports a set of estimation outcomes reflecting the main findings of the macro approach.

1 Introduction

This paper investigates how existing diasporas or networks (i.e. stock of immigrants of own national background already resident in a given destination) impact the number, skill composition and geographical concentration of new migrants. These diaspora externalities operate through two channels. On the one hand, diasporas reduce assimilation and information costs for newcomers. On the other hand, they attract new migrants through family reunion programs. Assessing the combined effect of these two channels is key to understand the dynamics of the size and composition of migration flows.

Until recently, the analysis of diaspora externalities has been conducted at the micro level (see Boyd, 1989, Massey, 1993, Munshi, 2003, Mc Kenzie and Rapoport, 2010). Microdata have multiple advantages. However a major drawback of those micro approaches is that they focus on a limited number of migration corridors (e.g. the Mexican-US corridor) and can hardly be generalized to other country pairs. An exception is provided by Beine et al (2010) who use bilateral macrodata on migration flows and stocks. They took advantage of a recent data set on international migration by educational attainment to investigate the role of diaspora size on the educational structure of migration from 195 countries to the 30 OECD countries. Their paper shows that networks are by far the most important determinant of migration flows, explaining 71 percent of the observed variability of the size of migration flows, and 47% of the variability of the selection ratio in 2000.

In this paper, we extend the study of Beine et al. by looking at the diaspora effect on the geographical concentration or dispersion of new migrants, and by comparing results obtained with different estimation techniques. Indeed, a macro analysis of diaspora externalities raises multiple econometric problems; the main issues are the large number of empty corridors (due to truncation rules or true 'zeroes'), and the difficulty to identify causation (unobserved variables are likely to affect the existing stock of migrants and the flows of newcomers). Several econometric techniques are available to address these issues; one of our goals is to evaluate the quantitative robustness of diaspora externalities to the choice of a particular method.

The remainder of this paper is organized as following. Section 2 depicts theory and key issues. Section 3 reviews the main empirical hurdles researchers face when studying the impact of diasporas. Section 4 discusses econometric issues while Section 5 presents some estimation results. Section 6 concludes.

2 Theory and key issues

To describe how existing diasporas impact the size, skill composition, and geographic concentration of migration flows, let us consider a simplified version of the model used in Beine et al. (2010). An individual endowed with h units of human capital earns a wage $w_i h$ in country $i = 1, ..., I$, where w_i is the skill price in that country. The skill price is linked to labor productivity and the

level of development of the country.

The utility of a type-h individual born in country o and staying in that country is given by: $\mathrm{u}_{oo}(h) = w_o h + A_o + \varepsilon_o$where A_o is a variable capturing non-wage income and amenities in the origin country.

The utility obtained when the same person migrates to country d is given by $\mathrm{u}_{od}(h) = w_d h + A_d - C_{od}(h) + \varepsilon_d$where $C_{od}(h)$ denotes moving and assimilation costs that are borne by the migrant, including visa fees.

Those costs depend on factors such as physical distance, destination and origin countries' social, cultural and linguistic characteristics. They are decreasing with human capital ($\partial C_{od}/\partial h < 0$), as high-skilled migrants are better informed than the low skilled, have higher capacity to assimilate or have more adaptive skills. In addition, visa costs are lower for high-skilled migrants when quality-selective programs exist in the destination country.

Let N_o denote the size of the native population that is within migration age in country o. When the random term ε_i follows an iid extreme-value distribution (see McFadden, 1974), the probability that a type-h individual born in country o will move to country d is given by: $\Pr\left[u_{od}(h) = k \max u_{ok}(h)\right] = \frac{\exp[w_d h + A_d - C_{od}(h)]}{\sum_k \exp[w_k h + A_k - C_{ok}(h)]}$,and the log-ratio of emigrants to country d to stayers is determined by

$$\ln\left[\frac{N_{od}(h)}{N_{oo}(h)}\right] = (w_d - w_o)\, h + (A_d - A_o) - C_{od}(h) \tag{1}$$

Albeit very simple, this one-skill model is rich enough to depict some key patterns of international migration:

- the size of bilateral migration flows increases with the wage differential $(w_d - w_o)$, differences in amenities $(A_d - A_o)$, and decreases with the level of migration cost (C_{od}).
- emigration rates are lower for high-skilled workers than for the high-skilled since the latter benefit more from wage differentials and have lower migration costs (positive selection).
- the proportion of high-skilled migrants is larger in countries with higher skill prices (positive sorting).

The model can also be used to understand how diasporas affects the magnitude and structure of migration flows. In what follows, we denote the bilateral diaspora size by M_{od}, show how diaspora externalities can be introduced in equation (1), and how they can be empirically estimated using bilateral data on migration stocks and flows.

2.1 Impact on size

Existing social networks reduce migration costs through two main channels. First, they lower information, assimilation and adaptation costs. Second, they

lower the visa costs by allowing migrants to benefit from family reunification programs. For these reasons, the diaspora size can be included in the determinants of migration costs: we write $C_{od}(h, M_{od})$ with $\partial C_{od}/\partial M_{od} < 0$.

The size-externality of diasporas can be tested by regressing skill-specific bilateral flows in logs, $\ln[N_{od}(h)]$, on the log of the stock of existing migrants at the beginning of the period, $\ln[M_{od}]$. Assuming a logarithmic functional for the diaspora effect, equation (1) can be rewritten as

$$\ln[N_{od}(h)] = \alpha_o^h + \alpha_d^h + \beta^h \ln[M_{od}] + \partial^h D_{od} + \eta_{od}^h \tag{2}$$

where $\alpha_o^h \equiv \ln[N_{oo}(h)] - w_0 h - A_0$ is an origin-country fixed effect, $\alpha_d^h \equiv w_d h + A_d$ is a destination country fixed effect, D_{od} is a vector of other observable distance variables affecting migration costs, and η_{od}^h is the error term. A positive value for β^h is expected if existing networks reduce migration costs.

2.2 Impact on selection

Positive selection is affected by diasporas if the effect of existing networks varies by skill group. Indeed, as assimilation and information costs are sources of positive selection (because they decrease with human capital), any cut in migration costs will be relatively more favourable to low-skilled migrants. Second, when the diaspora size is bigger, the probability that a migrant relies on an economic migration program declines and the probability she/he relies on family reunion programs increases. The advantages of being skilled are likely to be less important when migrants can rely on others.

Such an effect on positive selection can be indirectly evaluated assessed by differentiating the β^h obtained from the skill-specific regressions in (2). A more direct way to capture this externality is to regress the log-ratio of high-skilled to low-skilled migrants on the diaspora size. Denoting by $\overline{h}$ and $\underline{h}$ the human capital levels of high-skilled and low-skilled individuals, the dependent variable can be written as $\ln[S_{od}] \equiv \ln\left[N_{od}(\overline{h})/N_{od}(\underline{h})\right]$. From equation (1) we have

$$\ln[S_{od}] = \alpha_o + \alpha_d + \beta \ln[M_{od}] + \partial D_{od} + \eta_{od} \tag{3}$$

where $\alpha_o \equiv \ln\left[N_{oo}(\overline{h})/N_{oo}(\underline{h})\right] - w_0(\overline{h} - \underline{h})$ is a origin-country fixed effect, $\alpha_d \equiv w_d(\overline{h} - \underline{h})$ is a destination country fixed effect, and η_{od} is the error term. A negative value for β is expected if existing networks reduce positive selection.

2.3 Impact on relative concentration

We also investigate the question of the relative concentration of diasporas across education levels. In particular, we investigate whether diasporas lead to more concentration of low-skilled rather than high-skilled migrants. The impact of diasporas on the concentration level should be in line with the effect in terms of selection. In particular, if diasporas tends to favor a negative selection process, it should increase the concentration of low-skilled migrants compared to the concentration of high-skilled migrants.

Our relative concentration measure is equal to :

$$RC_{od} = \left[N_{od}(\overline{h}) / \sum_k N_{ok}(\overline{h}) \right]^2 - \left[N_{od}(\underline{h}) / \sum_k N_{ok}(\underline{h}) \right]^2 \quad (4)$$

A nice property of this bilateral measure is that its sum across destination countries boils down to the difference between Herfindhal indexes for high-skilled and low-skilled migrants.

The models for relative concentration writes as:

$$RC_{od} = \alpha_o + \alpha_d + \beta \ln [M_{od}] + \partial D_{od} + \eta_{od} \quad (5)$$

and a negative value for β is expected if existing networks increase the concentration of low-skilled migrants compared to the concentration of the high-skilled.

3 Key empirical issues

To illustrate, we will used bilateral migration macrodata to quantify the size of diaspora externalities, and evaluate the robustness of the elasticity to the choice of econometric technique. We use the Docquier, Lowell and Marfouk (2009, referred to as DLM from now on) database. Based on census and register information on the size and structure of immigration in all OECD countries, DLM database provides the stock of migrants from any given country to 30 OECD countries by education level in 1990 and 2000. The dataset covers only the adult population aged 25 and over, and migration is defined on the basis of the country of birth rather than citizenship[1]. It is worth noticing that the DLM database badly captures illegal immigration for which systematic statistics by education level and country of origin are not available. Some illegals are recorded in the US census. However, for the other member states of the OECD, data on illegal immigration are less reliable or do not exist. By disregarding illegal migrants, the database probably underestimates bilateral migration stocks/flows and overestimates the average level of education of the immigrant population.

The main strength of the DLM database is that it distinguishes between three levels of education for migrants. High-skilled migrants are those with post-secondary education. Medium-skilled migrants are those with upper-secondary education completed. Low-skilled migrants are those with less than upper-secondary education, including those with lower-secondary and primary education or those who did not go to school.

The main characteristics of the diaspora that we consider in this paper are the following:

[1] Even though this is the standard definition of a migrant, especially in the economics literature, the dataset does not include second generation children who are born in the destination country even though they might constitute an important part of a diaspora in the sociological sense. This is simply due to absence of comprehensive administrative data in tracking of the migrants' children. However, we expect diaspora sizes inclusive and exclusive of second generation to be highly correlated.

- the bilateral migration flow is proxied by the change in the stock of migrants born in country o and living in the OECD country d observed between 1990 and 2000 for each skill group.
- the bilateral indicator of positive selection is proxied by the log-ratio of the number of high-skilled to low-skilled new migrants from o to d (we disregard medium-skilled migrants)
- the bilateral indicator of relative concentration is the 'high-skill minus low-skill' difference in the squared proportions of migrants from a given origin country o to the 30 possible destinations, following equation (4).
- the size of the existing diaspora is measured as the immigrant population born in country i and living in the OECD country j ($\neq i$).

Of course the use of macro data is not without problems. In this section, we discuss the pros and cons of using macrodata, and the main issues triggered by a macro approach.

3.1 Micro vs macro approach

A first important distinction in the empirical approaches concerns the use of a microeconomic approach as opposed to a macroeconomic one. A large body of the literature use microeconomic data to study the impact of networks on the propensity to migrate and on the impact on the educational quality of the migrants (Massey, 1986; Munshi, 2003, Mc Kenzie and Rapoport, 2010). This contrasts with more recent approaches using macroeconomic data of international migration flows. Both approaches have their advantages and drawbacks and should see as complementary to address the key issues at stake here.

Micro approaches obviously yield interesting insights. By focusing on individuals or households, those approaches can account for the role of individual characteristics of the migrants in a straightforward way. One of those characteristics is the education level of the migrants, as shown by Mc Kenzie and Rapoport (2010). Their approach confirms that the sensitivity to network is much higher for uneducated migrants than for educated ones. Another appealing feature of the micro datasets is that they can distinguish between different levels of the network. As Massey (1986) proposes, the network effects can be decomposed as strong and weak ties between the migrant and the community at destination. Some data makes it possible to make a distinction between networks defined at the community level and networks at the household (family) level. The latter can be expected to provide different kinds of assistance such as financial aid and support. A final appealing element is the possibility of finding suitable instruments for the network at destination. This is desirable given the special statistical challenges at stake in this literature (see section on the reflection problem). Those instruments are supposed to be strong predictors of the network but uncorrelated with the dependent variable that is impacted by the network such as the size of the flows of their composition. A good example is

provided by Munshi (2003) for networks defined at the community level. Rain fall in origin communities in Mexico are supposed to predict the rate of emigration of those migrants but are uncorrelated with labor market outcomes at destination (in the US) that are potentially affected by Mexican immigration.

A major drawback of micro approaches is that they often consider only a limited number of migration corridors. A good example is provided by the Mexican-US migration patterns. In that particular case, the pitfalls are quite limited since the US is the destination for more than 99% of the Mexican migrants. This is nevertheless true for a large number of origin countries. Depending on their size, origin countries have diversified destination for their migrant communities. Even in origin countries characterized by a limited number of destinations, the patterns are likely to be very different across destinations. A good example is provided by Cape Verde, which sends a majority of unskilled migrants to Portugal and Luxembourg while sending most skilled migrants to the US. The possibility of pooling a large set of origin and destination countries in macro datasets makes it possible to assess statistically the determinants of the various patterns in international migration. Also, the informational content of the empty cells of migration flows in country pairs is valuable, since the zeros reflect that the net gains of migration in that corridor are too low for all natives. In turn, the low level of the net gains might reflect that some factors lead to high levels of migration costs, including the absence of a network at destination. In other terms, while it creates additional statistical issues, the inclusion of zero values in macro datasets tends to lower the case for selection biases.

3.2 The zeroes

Starting from the Docquier and Marfouk (2006) dataset including almost all origin countries and 30 destination countries, we can characterize some of statistical properties of the migration flows. The distribution of the migration flows turns out to be unimodal, highly left skewed with a large amount of zero values for both the (net) migration flows between 1990 and 2000 and the stocks in 1990 and 2000. For the flows, we have 34% of pairs of countries with zero values. For a restricted set of destination countries and for stocks, Grogger and Hanson (2010) report to have a bout 13% of zero values.

It would be nice to plot a kernel density estimation of the distribution of the (bilateral) diasporas

What do these zero values reflect? This is a tricky issue and the amount of information usable to discover their nature is very limited. For a small set of country pairs, the zero values might be the result of a truncation process. For instance, for reasons of statistical confidentiality, some low number of migrants of country i in country j. This is reported to be the case for provincial data of international migrants in Canada (see Head and Ries, 2003). Under 5 recorded migrants, the statistical offices are expected to repot a zero to preserve the anonymity of the migrants. Also, it is possible that a number diplomates are not counted in the official stock of migrants following international conventions.

More likely, a large number of zero values reflect true zeroes. Like in trade

data, some bilateral migration patterns are not 'profitable' so that there are simply no migrant to record. Ignoring those zero values would be highly detrimental to assess the relevance of the determinants in international migration since they reflect that the costs of migration might be too high for all potential migrants coming form country i. Among those factors, the fact that there is no network at destination might lead to cost levels that deter all migrants to choose that particular destination.[2] Therefore, it is important in the empirical investigation of the network effect to choose for methods that integrate those zero migration flows. In particular, for teh size , possible methods include Poisson regressions, 2-step Heckman approach and Tobit among others. For the selection and relative concentration however, Tobit and Poisson regression methods are not possible.

3.3 Stocks vs Flows

An important choice to be made in macroeconomic investigations of the network effect is the choice of the dependent variables. For instance, in their investigation of the determinants of international migration and in particular the role of the wage differential, Grogger and Hanson (2010) use stock data (observed in 2000). This allows to look at the long run effects of factors. Not surprisingly, variables such as colonial links turn out to be strong predictors of stocks in the long run. Colonial links exert two separate effects. They allow people to move during colonial times and shortly after independence. Part of those migrants are still included in the contemporaneous stock. a second more indirect effect is that colonial links create a dependance path for future migrants through the assimilation effect and family reunification.

To study the role of network in a very specific way, it is more desirable to use migration flows. The estimated relationship between migration flows over a specific period and the size of the stocks at the starting point of that period allows to capture the (short-run or medium-run) network effect. Interestingly, when colonial links are included in such a stock-flow model, they turn out to be insignificant. Their effect is absorbed by the network. From an economic point of view, the implications of those results is that recent migrants tend to come because they can rely on a network at destination, not because of past colonial links that offer little advantages nowadays.

The measurement of migration flows in destination countries is also a tricky issue. In most countries, we can rely on census data that provide the stock of migrants in a given year. For most census data, a ten-year frequency is the rule. Therefore, the only way to measure migration flows over a ten-year period is to take the difference between stocks. This in turn raises additional issues. First, the net migration flows are affected by the death rate of migrants included in the initial stock. Second, they are also affected by return migration. Return migration has recently received increased attention from the scientific

[2]Santos Silva and Tenreyro, 2006 show that ignoring the zero trade country pairs lead to some overestimation of some factors such as distance.

community. The rate of return migration is likely to be different across origin countries since the most important factors are also likely to be very different. For instance, using US data, Rosenzweig (2008) shows that the level of skill premia in the origin country is an important factor for the return migration rate of skilled migrants and students. Another issue is illegal migration and regularization programs implemented in destination countries. First, depending on the country, census data do or do not include illegal migrants in the stocks. For instance, the US census data include illegal migrants while most European data do not. Second, if regularization programs are implemented during the investigation period, this will raise the stocks of foreign born and in turn will increase the size of the migration flows. To sum up, we see that there are obviously negative and positive biases in measuring migrations flows through the change in stocks. Whether this tends to underestimate or overestimate the true values is obviously difficult to know. Another theoretical argument developed by Brcker(2006) suggests that using net migration flows instead of stocks might be misleading in case of heterogeneous agents. This is especially important when it comes to estimating teh impact of wage differentials on migration. In models like ours, we do not however explicitely include the wage differential for several reasons. One main reasons is teh absence of reliable data on wages in most origin countries. A second reason is that wage differential is the result of a combination of differences in base wage and differences in skill premia.

3.4 Defining a network

The macroeconomic investigation of the network effect relies on the the stock of nationals at destination. This is a natural definition of teh people supposed to provide assistance and help to teh new migrants. To what extent is that particular measure relevant. On the one hand, restricting teh diaspora to people with the same nationality might be restrictive. Ethnic networks are also known to be efficient and do not necessarily correspond to national borders. Migrants speaking the same language can be also very useful for the assimilation of new migrants. On the other hand, defining the network at the national level might overestimate the number of people able to provide help. Obviously, people located in large countries such as the US can provide assistance mainly to a restricted number of new migrants. This is especially true if concentration of migrants in the destination country is not very high.

3.5 The reflection problem

As explained by Manski (1993), one issue in identifying and estimating endogeneous social effects like the network effect here is the presence of unobservable correlated effects. In our framework, it could be the case that unobservable bilateral components will affect the size of the diaspora M_{ij} and the dependent variables. For instance, unobserved cultural proximity between country i and country j might affect simultaneously the stock of migrants, the current flows of new migrants and their selection. The cross-sectional nature of the data pre-

vents us to estimate directly those unobservable components. Therefore, those effects will be included in the error term, which in turn leads to some kind of omitted variable bias and to some correlation between M_{ij} and the error term.

4 Econometric Methods

There are several methods that can be used to estimate the impact of diasporas on flows, on their selection and on their relative concentration by skills.

A simple and easy way of estimation the models is OLS. Nevertheless, the high occurrence of zero values is likely to lead to inconsistent estimates. The use of a log specification drops the zero observations from the sample which is likely to result in biased estimates of the impact of diasporas and other variables on the migration flows and their selection. For instance, it might be the case that there are no migrants from country i to country j because migration costs are too high. In turn, migration costs might be too high because distance is too high and there is no diaspora. In this case, the exclusion of those observations leads to underestimation of the impact of the variables affecting the migration costs such as distance, colonial links, linguistic similarities or diasporas.

To minimize the bias due to selection issues, one possibility is to use Heckman 2-step estimation methods. In general, for all the features that we analyze (migration flows, selection and relative concentration), the first step involves the estimation of a selection equation - the probability for a given country pair to have a positive migration flow.[3] The usual procedure implies the use of an instrument in the probit equation, i.e. a bilateral variable that influences the probability of observing a diaspora between the two countries but does not influence the size of this diaspora.

It is difficult to find such an instrument but one possible candidate is diplomatic representation of the destination country in the origin country. Diplomatic representation might affect the probability of having at least one migrant by setting some kind of threshold on the initial migration and visa costs faced by potential migrants. In the absence of any diplomatic representation of country j in country i, the cost to get a visa can simply be too high so that nobody would consider to migrate to country j. The role of diplomatic representation in the migration process is to a certain extent analogous to the role played by a common religion for trade relationships. As argued by Helpman et al.(2007), a common religion (a proxy of costs of establishing business linkages) affects the extensive margin of trade (i.e. the probability of export) but not the intensive margin (i.e. trade volumes). In regressions (**??**-**??**), the use of a two-step Heckman approach yields intuitive results both for the flow and for the selection equation. In particular, for the selection equation, we find that diplomatic

[3]To be more precise, for the analysis of migration stock, the probability that a given observation will be included in the regression is directly related to the probability of observing a diaspora (either regardless of the skill level, either for a particular skill level) for this country pair. For the migration flows, the probability is exactly the same since we have no case of zero migration flow with positive values of the stock in 1990 and 2000. For the analysis of selection, the probability is related to the existence of a diaspora or at least a skilled diaspora.

representation of country j in county i tends to positively affect the probability of observing a diaspora of country i in country j. Furthermore, the mills ratio turns out to be significant in the flow equation, suggesting that accounting for a selection bias is important.

Since the observed level of diaspora in 1990 is used as a regressor, the use of diplomatic representation leads to some collinarity problems in the selection equation. In order to mitigate the collinearity problems, it is possible to run Heckman two-step regressions without any additional instrument. As stressed by Wooldridge (2002), the use of an additional instrument in the probit equation is not strictly necessary. The drawback of not using an additional instrument is that the Mills ratio might become highly collinear with the explanatory variables of the flow equation, which in turn lowers the significance of the coefficients. This is not the case for most of our regressions.

Another alternative to the OLS regressions is to use Poisson regression models that relies on pseudo maximum likelihood estimates, as advocated by Santos Silva and Tenreyro (2006) who show that the use of log linearization for gravity models leads to inconsistent estimates of the coefficients (such as the one relative to distance). A first reason, as mentioned before, is the exclusion of zero observations for the dependent variable. A second reason is that the expected value of the error will depend on the covariates of the model and hence will lead to estimation biases of the coefficient. The Poisson solution is nevertheless unfeasible for the selection and the concentration analyses. For the selection, the existence of zero values for $M_{i,j}(h)$ leads to undefined values for S_{ij}, which cannot be handled by the Poisson approach.

The above mentionned estimation methods do not address one solution proposed by Munshi (2003) is to estimate by IV the effects of M_{ij}. To that aim, one has to find instruments of M_{ij}, i.e. variables supposed to be uncorrelated with the flows but that are good predictors of the stocks. Beine et al. (2010) use two instruments for that sake. The first instrument is a dummy variable capturing whether the two countries were subject to a temporary guest worker agreement in the 60's and 70's. One can expect those guest worker agreements to exert a strong impact on the initial formation of a stock of migrants in the 60's and the 70's, hence influencing the stock in 1990. In contrast, it is unclear why those initial agreements would influence the contemporaneous migration flows beyond the impact exerted by the diaspora it-self. For instance, it turns out that guest worker agreements did not create any preferential treatment at the level of country pairs in the migration policy. Therefore, it is expected that these guest worker agreements are not themselves correlated with the bilateral unobservable components. Examples of such a process are illustrated for instance by the impact of the post-war guest worker agreements between Belgium and Italy or Spain.

The second instrument proposed by Beine et al. (2010) is a variable capturing the unobserved diaspora in the 60's through a combination of variables representing some push factor in country i , size in country i, openess and size in country j and distance between i and j. We use four different measures. The basic measure is $ln(pop_i * immst_j/dist_{ij}) * armedconflict_i$ where pop_i is

the population size in the 60's of country i, $immst_j$ is the immigration stock of country j in the 60's, $dist_{ij}$ is the distance between i and j and $armedconflict_i$ is a dummy variable capturing the occurrence of armed conflicts in country i during the 60's. To capture push-factors leading to emigration in the 1950s and 1960s, we only consider conflicts observed between 1946 and 1960. We distinguish minor conflicts (number of battle-related deaths between 25 and 999) denoted CONFL1 and wars (at least 1,000 battle-related deaths in a given year) denoted CONFL2. The variables CONFL1 and CONFL2 sum up the number of annual conflicts over the period 1946-1960.

IV estimation methods are suited to address all issues (Size, selection and concentration). Nevertheless, like OLS, they are subject to issues related to the selection bias. [4]

5 Results

Table 1 reports the results for five different estimation techniques : OLS (using lows as the dependent variable), heckman two stage method with and without an instrument for the selection, Maximum likelihood Poisson and IV regression (on the flows as well) using the two above mentioned instruments.

The results illustrate the strong robustness of the estimation of the key elasticity. This elasticity ranges between 0.62 and 0.76. It means that a 10% increase in the size of the network tends to increase the subsequent flows over the next ten years by about 7%. This result is in line with some previous results on the US literature. For instance, focusing only on family reunification, Jasso and Rosenzweig (1986, 198w) show that the multiplier associated to sponsored migration is about 1.2. If this is true for other countries, this means that our results suggest that the multiplier associated to the pure network effect (assimilation effect) should be around 1.5 for the US. This seems reasonable.

[4] A combination of Poisson regression models along with IV estimation is proposed by Tenereyro (2009) within the GMM framework. This is nevertheless beyond the scope of this paper. Furthermore, this is relevant only for the size issues.

Table 1. Determinants of migration flows

	(1) OLS	(2) Heck with	(3) Heck w/o	(4) Poisson	(5) IV
Lagged diasp	0.620	0.660	0.699	0.703	0.761
	(34.35)***	(47.97)***	(43.91)***	(16.20)***	(10.92)***
Col links	0.331	0.219	0.127	-0.312	-0.051
	(2.45)**	(2.03)**	(1.10)	(1.65)*	(0.26)
language	0.388	0.477	0.496	0.298	0.234
	(5.20)***	(6.71)***	(6.48)***	(2.53)**	(2.27)**
Log(dist)	-0.408	-0.501	-0.448	-0.337	-0.259
	(9.04)***	(12.04)**	(10.69)***	(3.28)***	(2.84)***
Schengen	0.168	0.257	0.277	0.061	0.160
	(1.19)	(2.00)	(2.02)**	(0.23)	(1.11)
Constant	3.750	2.785	2.365	3.461	2.365
	(6.92)***	(4.82)***	(4.02)***	(3.06)***	(2.69)
Observations	3608	5610	5760	5374	3486
Mills ratio	-	0.908	1.19	-	-
		(7.60)***	(9.35)***	-	-

Table 2. Impact of diaspora on selection (level and change in log high-skill/low-skill ratio)

	Log-skill ratio (OLS)	Log-skill ratio (Heck)	ΔLSR (Heck)	ΔLSR (IV)
Lagged diasp	-0.171	-0.194	-0.212	-0.215
	(16.19)***	(20.62)***	(17.62)***	(2.95)***
Col links	-0.042	-0.022	0.101	0.270
	(0.62)	(0.32)	(1.67)*	(1.77)*
language	0.466	0.460	0.176	0.235
	(9.38)***	(9.37)***	(4.17)***	(3.19)***
Log(dist)	0.096	0.090	0.086	0.019
	(3.35)***	(3.40)***	(3.78)***	(0.30)
Schengen	0.502	0.519	0.390	0.414
	(5.65)***	(6.26)***	(5.48)***	(6.08)***
Constant	-1.109	-0.734	-1.250	-0.481
	(1.16)	(1.32)	(2.54)**	(0.63)
Mills	-	(-0.380)	(-0.10)	-
	-	(6.86)***	(0.22)	-
F-stat First stage	-	-	-	30.07
Hansen J-test (p-value)	-	-	-	0.747
Observations	3486	5760	5760	3486

Absolute values of robust t statistics in parentheses

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

Instrument sets for M_{ij} in all columns include a dummy for bilateral guest-worker agreements and a proxy for diaspora size in 1960. In column (1) , the proxy is computed as $ln(pop_i * immst_j/dist_{ij}) * Conf1_i$.In column (2) , the proxy is computed as $ln(pop_i * immst_j/dist_{ij}) * Conf2_i$;in column (3), the proxy is computed as $ln(pop_i * immst_j/dist_{ij}) * (conf1_i + Conf2_i)$.

Table 3. Explaining relative concentration between high-skill and low-skill and change in relative concentration

	Rel conc (OLS)	Rel conc (Heck)	ΔRC (Heck)
Lagged diasp	-0.502	-0.514	-0.008
	(5.87)***	(9.67)***	(16.05)***
Col. links	-4.635	-4.619	-0.040
	(4.68)***	(10.69)***	(9.93)***
Language	0.338	0.321	-0.004
	(0.84)	(1.09)	(1.58)
Log(dist)	0.266	0.269	0.006
	(1.24)	(1.69)*	(3.78)***
Schengen	-0.193	-0.180	0.002
	(0.50)	(0.36)	(0.49)
Constant	5.607	-3.240	-0.037
	(0.29)	(1.19)	(1.60)
Mills		-0.405	-0.873
		(1.07)	(2.44)**
Observations	3920	5730	5730

Robust t statistics in parentheses

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

Table 2 looks at the selection, using the skill ratio (skilled migrants over unskilled migrants) as an indicator. Four different estimation results are reported. The first one uses OLS applied to the log of the ratio (observed in 2000). The second column reports the same estimate but with Heckman two stage method (without instrument). The third column does the same but on the change in the (log of) the skill ratio between 2000 and 1990. Finally, the last column also looks at the variation but using instrumental variable. The results shows that the network exerts important effect in terms of negative selection. The results hold for the level but also for the change in the skill ratio, regardless of the adopted method.

Table 3 investigates the same analysis but on the relative concentration between skilled and unskilled migrants. We use OLS, Heckman two stage method on the level and the change in the concentration measure. The results show that the networks tend to increase the concentration of unskilled migrants with respect to the skilled ones. This result is in line with the one concerning the selection of the migrants.

6 Conclusion

This paper reviews the existing literature on the impact of migrants networks on the patterns of international migration. It covers the theoretical channels at stake in the global effect of the networks. It identifies the key issues, namely the impact on size, selection and concentration of the migration flows. The paper also reviews the empirical hurdles that the researchers face in assessing the importance of networks. The key issues concern the choice of micro vs a macro approach, the definition of a network, the access to suitable data and the adoption of econometric methods accounting for the main features of those data.

The main results of the macro approach are illustrated with estimation results obtained with the Docquier-Marfouk (2006). It turns out that the networks exert strong influence on the international flows as well as a negative selection effect. They also favour the concentration of the unskilled migrants with respect to the skilled ones.

7 References

Bertolini, S. (2009), "Networks, Sorting and Self-selection of Ecuadorian Migrants", Paper presented at the second TOM Meeting, Louvain-La-Neuve, January.

Borjas, G (1987), "Self-selection and the earnings of migrants", American Economic Review, 77 (4), 531-53.

Borjas, G.J. (1994), "The economics of immigration", Journal of Economic Literature, 32, 1667-1717.

Borjas, G.J. (1995), "The economic benefits from immigration", Journal of Economic Perspectives, 9 (2), 3-22.

Borjas, G.J. (1999), Heaven's door: immigration policy and the American economy, Princeton University Press.

Carrington, W.J., E. Detragiache and T. Vishwanath (1996), "Migration with endogenous moving costs", American Economic Review, 86 (4), 909-30.

Chiquiar, D. and G.H. Hanson (2005), "International migration, self-selection, and the distribution of wages: evidence from Mexico and the United States", Journal of Political Economy, 113 (2), 239-81.

Clair, G., G. Gaullier, Th. Mayer and S. Zignago (2004), "A note on CEPII's distances measures", Explanatory note, CEPII, Paris.

Cohen, A. and A. Razin (2008), "Skill composition of migration and the generosity of the welfare state: free vs. policy-restricted migration", Mimeo., Tel-Aviv University.

Docquier, F. and E. Lodigiani (2009), "International migration and business networks", Open Economies Review, forthcoming.

Docquier, F., O. Lohest and A. Marfouk (2007), "Brain drain in developing countries", World Bank Economic Review, 21, 193-218.

Docquier, F. and A. Marfouk (2006), "International migration by educational attainment (1990-2000)", in C. Ozden and M. Schiff (eds). International Migration, Remittances and Development, Palgrave Macmillan: New York (2006), chapter 5.

Docquier, F., B.L. Lowell and A. Marfouk (2007), "A gendered assessment of highly skilled emigration", Population and Development Review, 35 (2), 297-321.

Friedberg, R.M. and J. Hunt (1995), "The impact of immigrants on the host country wages, employment and growth", Journal of Economic Perspectives, 9, 23-44.

Gao, T.(2003), "Ethnic Chinese Networks and International Investment: Evidence from Inward FDI in China", Journal of Asian Economics, 14, 611-629.

Gleditsch, P., M. Eriksson and M. Sollenberg (2002), "Armed Conflict 1946-2001: A New Dataset", Journal of Peace Research, 39 (5), 615-637.

Grogger, J and G.H. Hanson, 2008, "Income Maximisation and the selection and sorting of international Migrants, NBER Working Paper, No. 13821.

Harbom, L., E. Melander and P. Wallensteen (2007), "Dyadic Dimensions of Armed Conflict, 1946—2007", Journal of Peace Research, 45 (5), 697-710.

Helpman, E., M. Melitz and Y. Rubinstein (2007), "Estimating Trade Flows: Trading Partners and Trading Volumes", NBER Working Paper W12927.

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

Massey, D.S., J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino and J. E. Taylor (1993), "Theories of international migration: Review and Appraisal," Population and Development Review, 19 (3), 431-466.

McFadden, D. (1984), "Econometric analysis of qualitative response models", in: Z. Griliches and M. Intriligator, eds., Handbook of Econometrics, Volume 2, Amsterdam. Elsevier/North-Holland.

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

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

Pedersen, P.J., M. Pytlikova and N. Smith (2008),. "Selection and network effects–Migration flows into OECD countries 1990-2000", European Economic Review, 52 (7), 1160-1186.

Rapoport, H. and M. Kugler (2006), "Skilled Emigration, Business Networks and Foreign Direct Investment", CESifo Working Paper Series No. 1455.

Rauch, J. (2003), "Diasporas and development: Theory, Evidence and Programmatic Implications", Department of Economics, University of California at San Diego.

Rauch, J. and A. Casella (1998), "Anonymous Market and Group ties in International Trade", Journal of International Economics, vol 58(1):19-47.

Rauch, J. and V. Trindade (2002), "Ethnic Chinese Networks In International Trade", The Review of Economics and Statistics, MIT Press, vol. 84(1):116-130.

Razin, A. and E. Sadka (2004), " Welfare migration: Is the net fiscal burden a good measure of its economic impact on the welfare of the native-born population?", NBER Working Paper 10682.

Rosenzweig, M (2008), The global Migration of Skill, Paper presented at the Migration and Development Workshop, Lille, June.

Roy, A.D. (1951), "Some thoughts on the distribution of earnings", Oxford Economic Papers, 3 (2), 135-46.

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

Williamson, J.G. (2006), "Global migration: Two centuries of mass migration offers insights into the future of global movements of people", Finance and Development, 43 (3).

Wooldridge, J.M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press.