Determinants of the brain drain

F.D.

UCL

June 2008

Structure of the talk

1. **Introduction**
2. Determinants of the brain drain
3. Size and structure of diasporas

Motivation

- BD usually seen as a cause of poverty but, it can also be viewed as a consequence of poverty. Possibility of vicious circles: poverty induces skilled migration, skilled migration exacerbates poverty.
- Many host countries want to reinforce the selection of their immigrants. Is it easy and feasible?

Empirically, for sending and host countries, it is important to understand what causes the BD:

- What determines the desire to leave?
- What determines the size and structure of migration flows?

Sources

- Docquier F., O. Lohest, A. Marfouk (2007), Brain drain in developing countries, World Bank Economic Review 21 (2007), 193-218,
- Beine M., F. Docquier and C. Ozden (2008), Diasporas, work in progress

Structure of the talk

1. Introduction
2. **Determinants of the brain drain**
3. Size and structure of diasporas

Determinants - sources

- Data: Emigration rates computed by Docquier-Marfouk (2006)
- Emigration rates: $m_{i,t}^k = \frac{M_{i,t}^k}{N_{i,t}^k + M_{i,t}^k}$

TABLE 1. Descriptive Statistics by Country Group, 1990–2000

Group of origin	Emigration structure			Skilled emigrants by destination			Labor force structure (region of origin)			Emigration rates	
	Total emigrants (ages 25 and older, thousands)	Skilled emigrants (ages 25 and older, thousands)	Share of skilled (%)	In selective-immigration countries (%)	In EU15 countries (%)	In rest of OECD (%)	Total labor force (ages 25 and older, thousands)	Skilled labor force (ages 25 and older, thousands)	Share of skilled (%)	Total (%)	Skilled (%)
2000											
World[a]	59,022	20,403	*35*	73	21	6	3,187,233	360,614	11	1.8	5.4
High-income countries	19,206	7,547	*39*	68	24	8	666,246	200,607	*30*	2.8	3.6
Developing countries	38,083	12,576	*33*	76	19	5	2,520,987	160,008	*6*	1.5	7.3
Low-income countries	6,544	2,948	*45*	77	21	1	898,768	36,332	*4*	0.7	7.5
Lower-medium-income countries	17,053	6,089	*36*	77	17	6	1,298,233	76,981	*6*	1.3	7.3
Upper-medium-income countries	14,486	3,539	*24*	75	20	5	323,987	46,694	*14*	4.3	7.0
Least developed countries	2,510	853	*34*	69	29	2	245,974	5,635	*2*	1.0	13.1
Landlocked developing countries	1,271	470	*37*	63	33	4	129,988	8,892	*7*	1.0	5.0
Small developing islands	4,001	1,504	*38*	90	9	1	24,979	2,041	*8*	13.8	42.4
Large developing countries (>40 million)	19,828	6,926	*35*	82	13	5	2,050,014	117,433	*6*	1.0	5.6

(Continued)

TABLE 1. *Continued*

Group of origin	Emigration structure			Skilled emigrants by destination			Labor force structure (region of origin)			Emigration rates	
	Total emigrants (ages 25 and older, thousands)	Skilled emigrants (ages 25 and older, thousands)	Share of skilled (%)	In selective-immigration countries (%)	In EU15 countries (%)	In rest of OECD (%)	Total labor force (ages 25 and older, thousands)	Skilled labor force (ages 25 and older, thousands)	Share of skilled (%)	Total (%)	Skilled (%)
1990											
World[a]	41,845	12,462	*30*	76	17	7	2,369,431	209,225	*9*	1.6	5.0
High-income countries	18,165	5,613	*31*	74	17	9	586,069	139,458	*24*	3.0	3.9
Developing countries	19,402	5,804	*30*	79	17	4	1,783,362	69,767	*4*	1.1	7.7
Low-income countries	3,454	1,267	*37*	77	21	1	677,539	21,291	*3*	0.5	5.6
Lower-medium-income countries	8,740	2,883	*33*	81	14	5	938,974	34,948	*4*	0.9	7.6
Upper-medium-income countries	7,208	1,654	*23*	77	19	4	166,848	13,528	*8*	4.1	10.9
Least developed countries	1,384	373	*27*	70	29	2	185,034	3,092	*2*	0.7	10.8
Landlocked developing countries	444	150	*34*	69	29	3	73,330	1,613	*2*	0.6	8.5
Small developing islands	2,595	866	*33*	91	9	1	19,371	1,059	*5*	11.8	45.0
Large developing countries (>40 million)	9,312	2,890	*31*	83	13	4	1,430,178	50,707	*4*	0.6	5.4

[a]Sum of emigrants from high-income countries, developing countries, and dependent territories and emigrants who did not report their country of birth.

Source: Docquier and Marfouk 2006.

Determinants - Ten facts

1. In 2000 developing countries accounted for 64.5 percent of total immigrants and 61.6 percent of skilled immigrants in the OECD
2. Skilled immigation: the share of developing countries in 2000 is 15 percentage points higher than in 1990.
3. About 3/4 of these immigrants live in one of the 3 most important host countries with selective-immigration policies.
4. One-fifth of them live in the EU15
5. Locations choices are linked to geographic distances, historical ties (are they linked to economic and political variables?)
6. The proportion of skilled workers among migrants (on average 33 percent for developing countries) is much higher than the proportion among residents (on average 6 percent). Hence, skilled emigration rates (on average 7.3 percent) are much higher than average emigration rates (on average 1.5 percent).

Determinants - Ten facts

7. Between 1990 and 2000 the average emigration rate of developing counries rose from 1.1 to 1.5 percent.
8. Although the proportion of skilled migrants increased, the skilled emigration rate decreased from 7.7 to 7.3 percent as the general level of schooling increased in developing countries.
9. The highest brain drain rates are observed in small island developing countries and in the least developed countries, and the lowest rates in large and landlocked developing countries.
10. Setting aside small island economies, the highest average brain drain rates are observed in Sub-Saharan Africa (13 percent), Latin America and the Caribbean (11 percent), and the Middle East and North Africa (10 percent).

Determinants - multiplicative decomposition

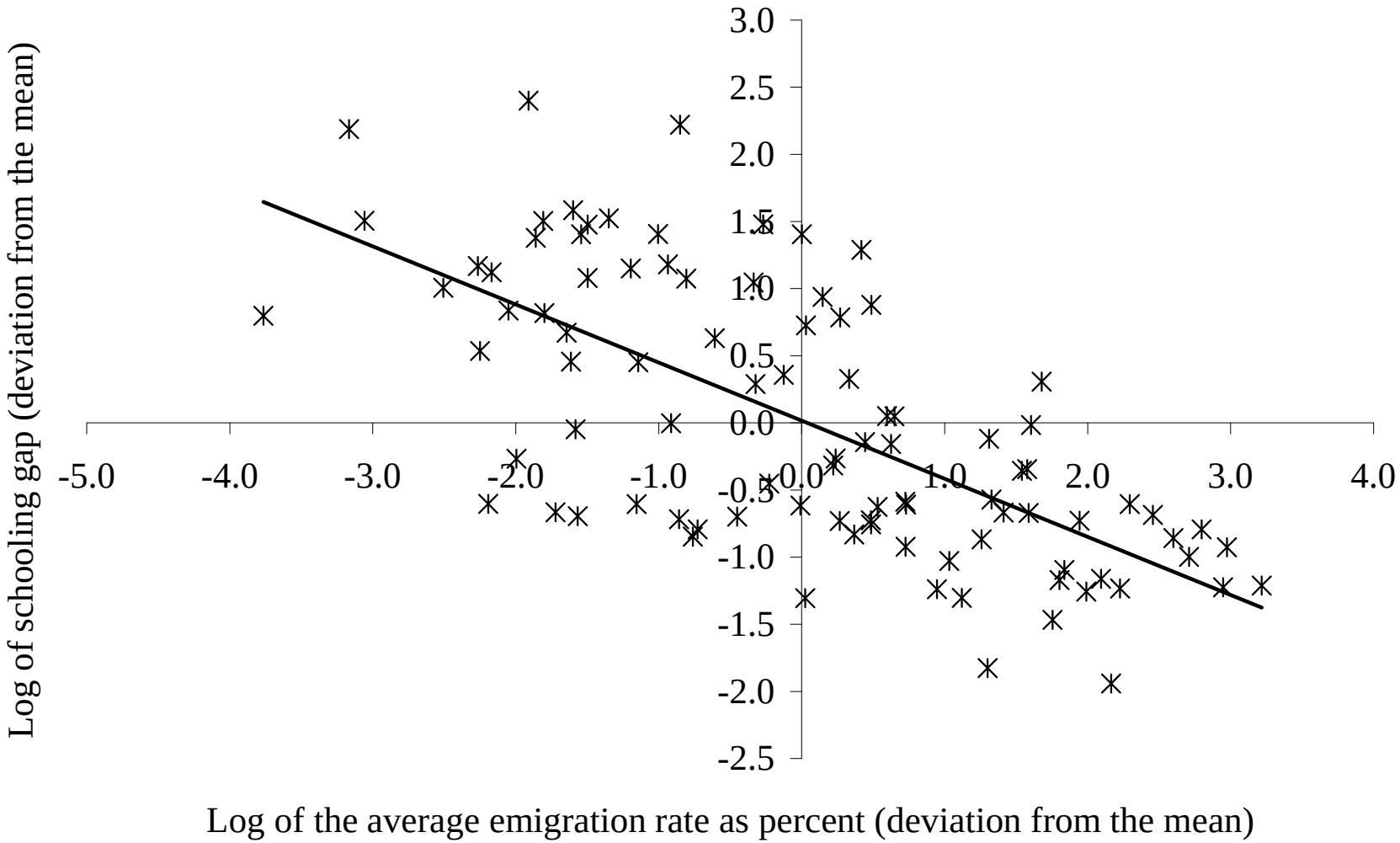
- A simple multiplicative decomposition of the skilled emigration rate can help to explain the distribution of the brain drain across countries:

$$m_{i,t}^h = \frac{M_{i,t}^h}{N_{i,t}^h + M_{i,t}^h} = \left(\frac{\sum_s M_{i,t}^s}{\sum_s \left[N_{i,t}^s + M_{i,t}^s \right]} \right) \cdot \left(\frac{M_{i,t}^h}{\sum_s M_{i,t}^s} \Big/ \frac{N_{i,t}^h + M_{i,t}^h}{\sum_s \left[N_{i,t}^s + M_{i,t}^s \right]} \right)$$

- First component = average emigration rate (ratio of emigrants to natives), reflects the degree of openness
- Second component = schooling gap (ratio of the share of skilled emigrants to the share of skilled natives)
- Hypothetical world where emigration is proportional to population and skill structure of emigration is identical to that of natives: BD would be homogeneous across countries
- However, many factors affect the degree of openness and the schooling gap.

Determinants - stylized facts 1

Average emigration rates and schooling gaps are negatively correlated

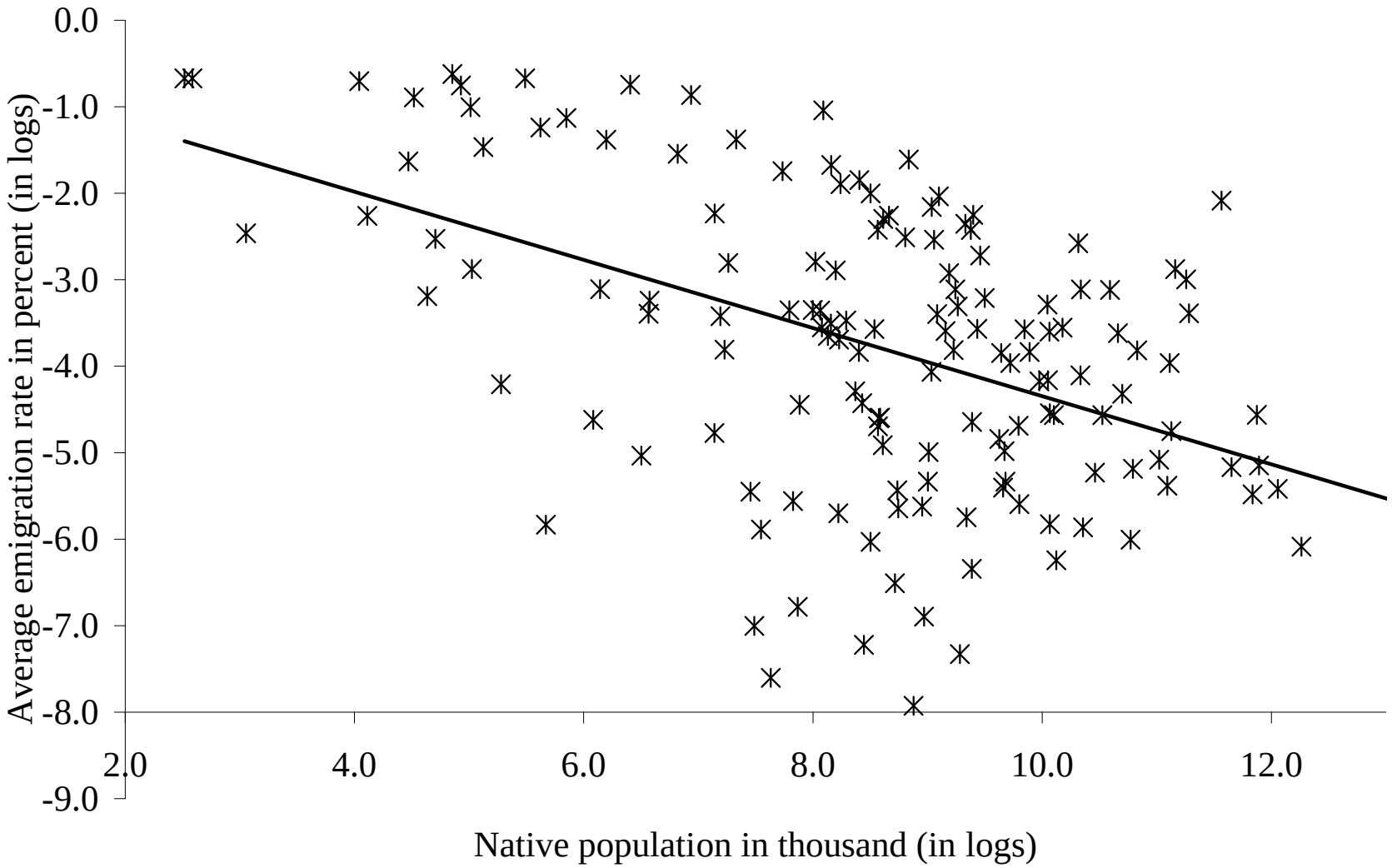


Determinants - stylized facts 1

- No developing country has both strong openness and a high schooling gap.
- If a country suffers from a large brain drain it is either because it is very open or because the positive selection of migrants is strong.
- This justifies the decomposition and the analysis of the specific determinants of these two components.

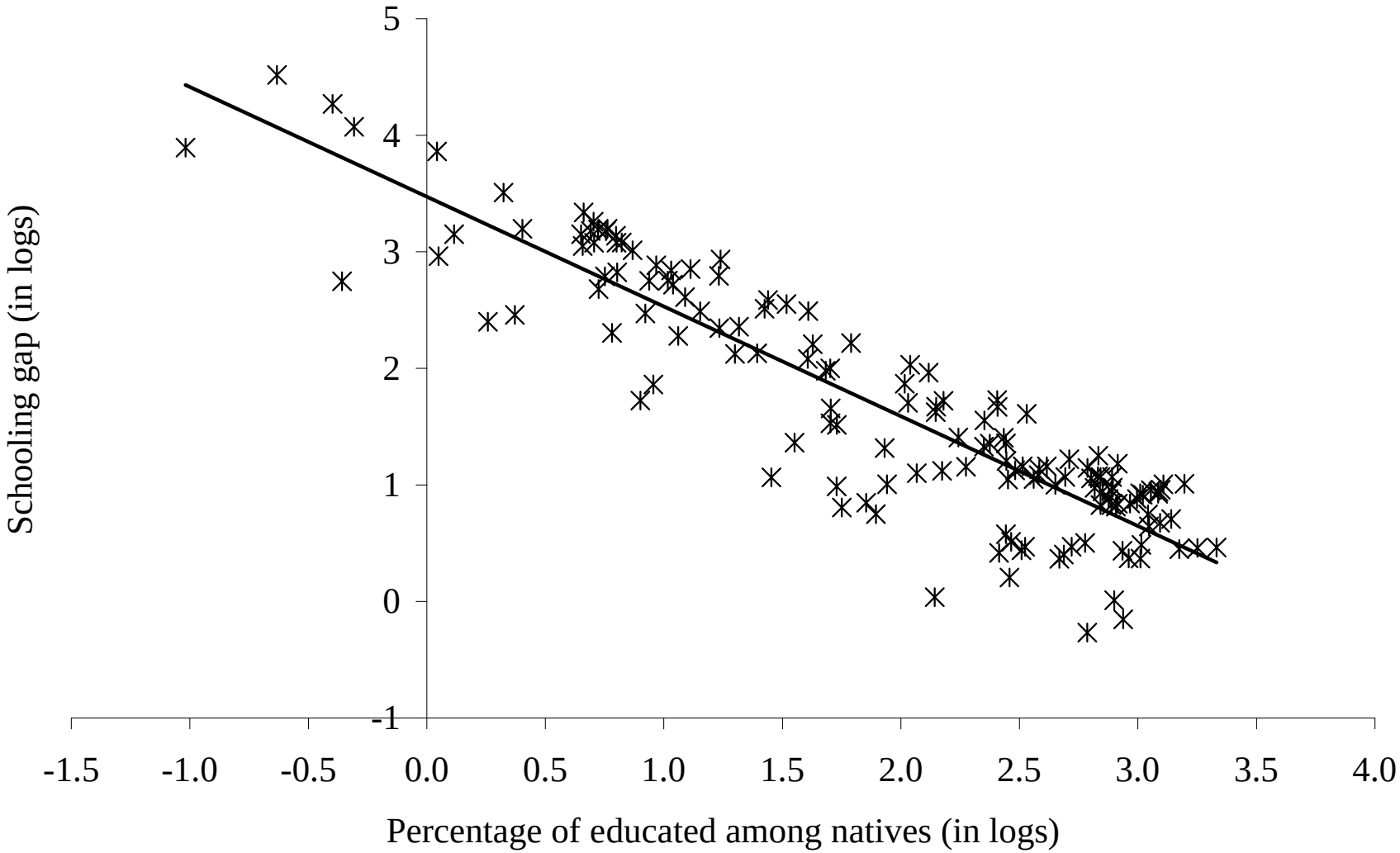
Determinants - stylized facts 2 (openess)

Average emigration rates decrease with country size ($corr = -.5$)



Determinants - stylized facts 3 (sch. gap)

Schooling gaps decrease with natives' rising human capital ($corr = -.9$)



Determinants - stylized facts 4 (sch. gap)

Schooling gaps depend on destination choice:

- On average, the schooling gap observed in selective-immigration countries was about twice as large as the gap observed in EU15 and other OECD countries in 2000.
- Countries that send many migrants to North America and Australia are likely to exhibit stronger schooling gaps than the others.
- Need of a bilateral analysis endogenizing location choices (see network analysis)

TABLE 2. Decomposition of Skilled Emigration Rates, 1990–2000

Group of origin	Decomposition			Openness by destination (%)			Schooling gap by destination		
	A	B	C						
	Brain drain (%) A = B × C	Openness (%)	Schooling gap	To selective-immigration countries	To EU15 countries	To rest of OECD	To selective-immigration countries	To EU15 countries	To rest of OECD
2000									
World[a]	5.3	1.8	3.0	1.0	0.6	0.2	3.8	1.9	1.9
High-income countries	3.6	2.8	1.3	1.4	1.1	0.3	1.7	0.8	0.9
Developing countries	7.3	1.5	4.9	0.9	0.5	0.1	6.1	3.0	3.0
Low-income countries	7.5	0.7	10.4	0.4	0.3	0.0	13.0	6.2	6.2
Lower medium-income countries	7.3	1.3	5.7	0.7	0.4	0.1	7.7	2.9	3.2
Upper-medium-income countries	7.0	4.3	1.7	2.8	1.2	0.3	1.9	1.2	1.2
Least developed countries	13.1	1.0	13.0	0.5	0.5	0.0	16.9	8.6	9.9
Landlocked developing countries	5.0	1.0	5.2	0.5	0.4	0.1	6.7	4.0	2.4
Small island developing economies	42.4	13.8	3.1	11.4	2.3	0.1	3.3	1.8	2.6

(Continued)

TABLE 2. *Continued*

Group of origin	Decomposition			Openness by destination (%)			Schooling gap by destination		
	A	B	C						
	Brain drain (%) A = B × C	Openness (%)	Schooling gap	To selective-immigration countries	To EU15 countries	To rest of OECD	To selective-immigration countries	To EU15 countries	To rest of OECD
Large developing countries (>40 million)	5.6	1.0	5.8	0.7	0.2	0.1	6.8	3.6	4.5
1990									
World[a]	5.2	1.6	3.3	0.9	0.5	0.1	4.6	1.7	2.4
High-income countries	3.9	3.0	1.3	1.6	1.0	0.4	1.8	0.7	0.9
Developing countries	7.7	1.1	7.1	0.6	0.4	0.1	9.7	3.7	5.0
Low-income countries	5.6	0.5	11.1	0.3	0.2	0.0	16.4	5.6	6.4
Lower medium-income countries	7.6	0.9	8.3	0.5	0.3	0.1	11.8	3.6	5.3
Upper-medium-income countries	10.9	4.1	2.6	2.7	1.3	0.2	3.2	1.7	2.4
Least developed countries	10.8	0.7	14.5	0.3	0.4	0.0	22.3	8.8	11.3
Landlocked developing countries	8.5	0.6	14.1	0.3	0.3	0.0	19.3	9.5	12.8
Small island developing economies	45.0	11.8	3.8	9.6	2.6	0.1	4.4	2.6	5.4
Large developing countries (>40 million)	5.4	0.6	8.3	0.4	0.2	0.0	10.6	3.9	7.0

[a]Sum of emigrants from high-income countries, developing countries, and dependent territories and emigrants who did not report their country of birth.

Source: Authors' analysis based on data from Docquier and Marfouk (2006).

Determinants - empirical analysis

Two-equation system
The dependent variables are:

- the logistic transformation of the average emigration rate. The dependent variable is ln[m/(1-m)], expands the range of the variable from (0,1) to (-inf,+inf).
- the log of the schooling gap.

Determinants - explanatory variables

- First set - country size (log of the native population, a dummy for SIDC)
- Second set - the level of development (log of the proportion of post-secondary-educated natives, log of GNI per capita in PPP, dummy for the LDC, dummy for oil exporting countries)
- Third set - sociopolitical environment at origin (political stability, government effectiveness, religious fractionalization)
- Fourth set - geographic and cultural proximity between developing and OECD countries (distance from selective-immigration countries, distance from the EU15, dummy for landlocked DC, dummy for former colonies of an OECD country, linguistic proximity, dummy if main destination is a selective-immigration country or EU15 member state)

Determinants - econometric issues

- Use of 2000 emigration rates, controls computed on 1980-2000
- 108 observations
- OLS regressions (same results with Random effects 90-00, pooling or SURE)
- White-correction for heteroskedasticity
- Endogeneity of human capital of natives (model IV, instruments = lagged HC, public education expenditures)

TABLE 3. Cross-Section Regression Results (2000 data)

	OLS-1 General model		OLS-2 Parsimonious model		IV-1 Parsimonious model		Larger sample model	
Variable	Openness[a]	Schooling gap[b]	Openness[a]	Schooling gap[b]	Openness[a]	Schooling gap[b]	Openness[a]	Schooling gap[b]
Country size								
Native population (logs)	−0.156 (1.79)*	0.019 (−0.58)	−0.178 (2.84)***		−0.173 (2.51)**		−0.153 (2.21)**	
Small island developing economies	0.779 (1.89)*	0.001 (0.00)	0.971 (2.90)***		1.013 (2.57)**		0.693 (1.81)*	
Level of development								
Proportion of post-secondary educated natives × 100 (logs)	0.744 (3.06)***	−0.883 (10.1)***	0.526 (4.05)***	−0.871 (11.4)***	0.663 (4.82)***	−0.795 (8.57)***	0.854 (5.01)***	−0.893 (14.8)***
GNI per capita (logs)	−0.129 −0.56	−0.144 (1.67)*		−0.091 −1.6		−0.135 (1.85)*		
Least developed country	−0.083 (−0.17)	−0.040 (−0.28)						
Oil exporting country	−0.650 (−1.57)	0.239 (1.81)*		0.161 (−1.23)		0.152 (−1.38)	−0.853 (2.67)***	0.188 (1.66)*
Sociopolitical environment								
Political stability	−0.082 (−0.39)	−0.002 (−0.03)					−0.300 (2.19)**	−0.061 (−1.66)*
Government effectiveness	0.007 (−0.03)	0.115 (−1.08)						
Religious fractionalization	0.376 (−0.83)	0.545 (3.06)***		0.578 (3.88)***		0.585 (4.05)***		0.509 (3.49)***
Geographic and cultural proximity								
Distance from selective-immigration countries (logs)	−1.143 (3.17)***	0.358 (2.35)**	−1.078 (3.01)***	0.445 (5.18)***	−0.924 (2.86)***	0.475 (5.09)***	−1.105 (3.82)***	0.479 (5.63)***
Distance from EU15 countries (logs)	−0.428 (3.23)***	0.113 (2.06)**	−0.389 (3.83)***	0.130 (2.39)**	−0.377 (2.96)***	0.139 (2.77)***	−0.398 (3.16)***	0.126 (2.37)**
Landlocked developing country	−0.872 (2.49)**	0.137 (−1.19)	−0.793 (2.37)**		−0.721 (2.51)**		−0.710 (2.47)**	
Former colony of an OECD country	0.318 (−1.00)	−0.024 (−0.22)					0.553 (2.12)**	
Main destination = selective-immigration countries	−0.001 (0.00)	0.757 (4.17)***		0.902 (5.89)***		0.920 (2.43)**		0.381 (3.80)***

(Continued)

TABLE 3. *Continued*

	OLS-1 General model		OLS-2 Parsimonious model		IV-1 Parsimonious model		Larger sample model	
Variable	Openness[a]	Schooling gap[b]	Openness[a]	Schooling gap[b]	Openness[a]	Schooling gap[b]	Openness[a]	Schooling gap[b]
Main destination = EU15	0.154 (−0.38)	0.403 (1.80)*		0.537 (3.01)***		0.614 (−1.59)		
Same language as a selective-immigration country	0.122 (−0.39)	0.154 (−1.63)						0.136 (1.80)*
Constant	11.672 (2.96)***	−0.794 −0.48	10.863 (3.31)***	−1.942 (1.89)*	9.052 (2.56)**	−2.100 (1.84)*	9.849 (2.93)***	−2.431 (2.38)**
Observations	108	108	108	108	108	108	125	123
Adjusted R-squared	0.67	0.88	0.68	0.88	0.69	0.89	0.68	0.89
Overidentification test[c]					0.12	0.13	0.33	0.88
Instrument relevance: *p*-value of *F* statistic					0.000	0.000	0.000	0.000
Exogeneity test[d]					0.07	0.26	0.07	0.27

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Numbers in parentheses are standard errors. Due to heteroskedasticity, the IV method for the schooling gap equation is a general method of moments estimator. Heteroskedastic-robust standard errors for OLS.

[a]Logistic transformation of the average emigration rate.

[b]Schooling gap in logs.

[c]*p*-value of statistic: Sargan test for the openness and Hansen *J* test for the schooling gap.

[d]Exogeneity test of natives of proportion skill. *p*-value of χ^2: Durbin-Wu-Hausman test for the openness and C-test for the schooling gap. List of instruments: lagged level + public expenditures in primary education (in logs).

Source: Authors' analysis based on data from Docquier and Marfouk (2006).

Determinants - results

- Country size - key determinant of openness but no effect on the schooling gap
- Level of development - very strong effect on openness rates and schooling gaps. The lower the natives' level of schooling, the greater is brain drain.
- Sociopolitical environment - significant impact. BD increases with religious fractionalization and political instability.
- Proximity significantly affects openness and the schooling gap.
- Countries that send most of their migrants to selective-immigration countries experience stronger schooling gaps.

Structure of the talk

1. Introduction
2. Determinants of the brain drain
3. **Size and structure of diasporas**

Diasporas: issues

- Diaspora: stock of migrants who gather in a given destination country.
- What are the driving forces affecting the size, skill composition and concentration of diasporas?
- Characterization of the dynamics of diasporas: do existing diasporas affect future migrations flows and their composition in terms of skills?

Diasporas: issues

"On the day I left Nigeria, I felt sad because I was leaving my family behind. I believed I would return eight years later, probably marry an Igbo girl, and then spend the rest of my life in Nigeria. But 25 years ago, I fell in love with an American girl, married her three years later, and became eligible to sponsor a Green Card visa for my 35 closest relatives, including my parents and all my siblings, nieces and nephews. The story of how I brought 35 people to the United States exemplifies how 10 million skilled people have emigrated out of Africa during the past 30 years. We came to the United States on student visas and then changed our status to become permanent residents and then naturalized citizens. Our new citizenship status helped us sponsor relatives, and also inspired our friends to immigrate here." (Philip Emeagwali)

Diasporas: issues

- Chain migration (*"inspired our friends"*) - Migration costs are endogeneous and depend on the size and the structure of the diaspora. Early evidence of Carrington, Detragiache and Vishwanath (1996) from US internal black migration: migrants'network tends to decrease costs for new migrants.
- Through family reunion programs (*"sponsor relatives"*), earlier cohorts of migrants attract future cohorts
- Our goal: study the dynamics of diasporas in size, structure and relative concentration

Diasporas: Stylized facts

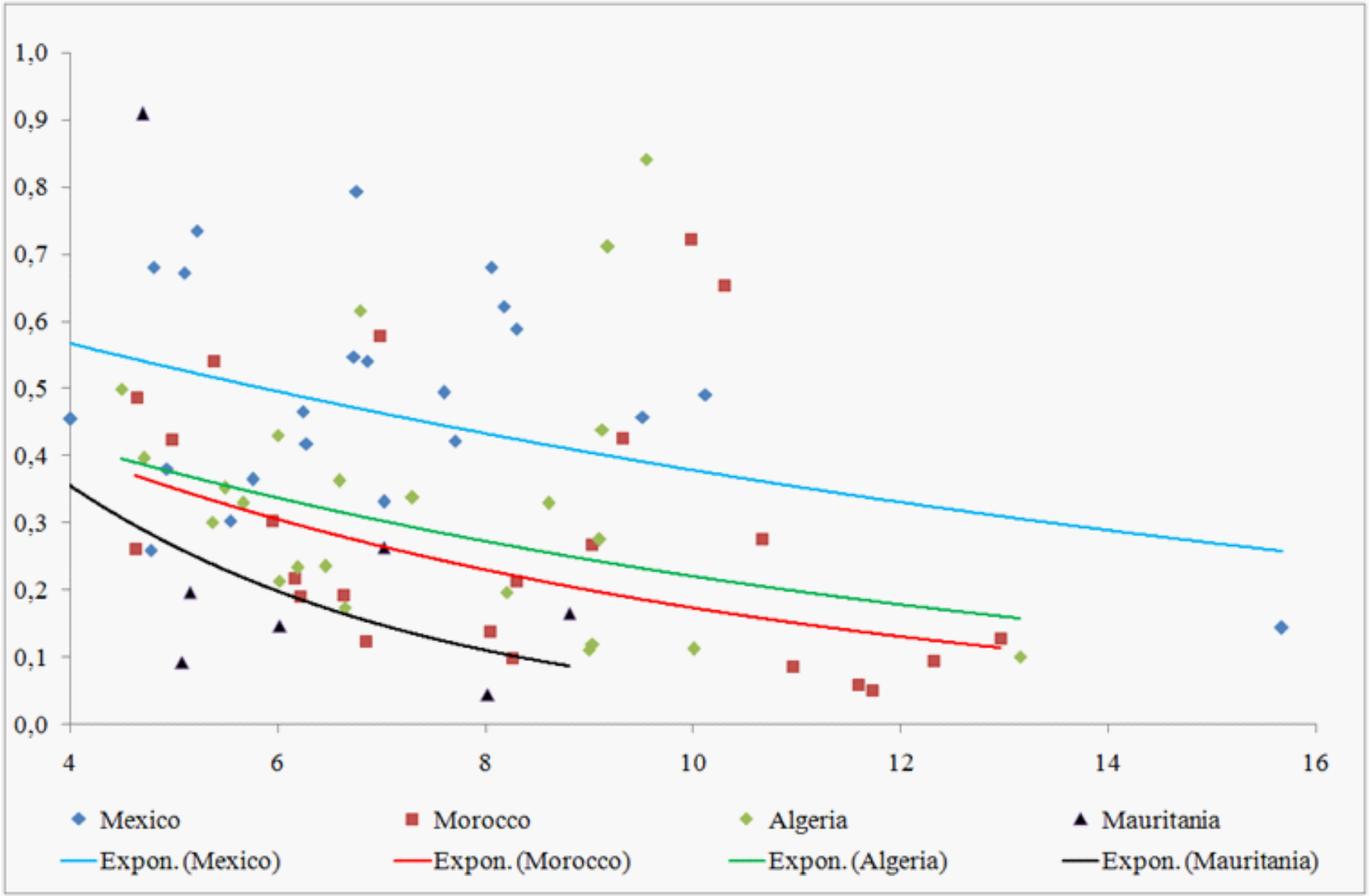
Important issue for host countries:

- Many factors affect the structure of the diaspora (distances, colonial links, wage differentials, immigration policies). Existing diasporas matter.
- Network effects: Turkey-Luxembourg : 44 pct skilled-26 pct unskilled), Turkey-Spain (33 pct skilled; 29 pct unskilled), Turkey-Germany : 6 pct skilled-86 pct unskilled
- Quantification of the impact of diaspora effects accounting for the other factors.

Diasporas: Stylized facts

- No clear relationship between diaspora size and its educational structure (either after dropping zeros or not).
- Nevertheless, at the bilateral level , we observe a negative relationship.

Diasporas: Stylized facts (bilateral)



MODELING THE DYNAMICS OF DIASPORAS

Diasporas: notations

- One-skill model with linear utility function (see Grogger and Hanson, 2008) and heterogeneous skills h: $U(h) = W(h)$, constant marginal utility of income.
- Country i's native population in age of emigration: N_i
- Expected wage of a worker with human capital h working in country j: $w_j h$ (variation in wages within countries is due to variation in skill levels; variation in average wages between countries is due to inter-country differences in average skill levels and skill prices; see Rosenzweig, 2008).
- Migration cost from i to j depends on macro and micro characteristics: distances ($d_{i,j}$), human capital (h), size of the diaspora abroad ($M_{i,j}$) and some home country's characteristics (x_i):

$$C_{i,j}(h) = c(d_{i,j}, h, M_{i,j}, x_i)$$

with $c_d^{'} \succ 0$, $c_h^{'} \leq 0$, $c_M^{'} \prec 0$ and $c_{Mh}^{''} \gtrless 0$.

Diasporas: a two-stage model

- The magnitude and structure of migration flows result from self-selection mechanisms (individual decisions) and out-selection mechanisms (immigration policies at destination)
- We consider a two-stage model. First, individuals decide whether or not to apply for a visa. Second, either they get it or not.
- Assumption: if they do not get their visa, they stay in their home country.

Diasporas: self-selection stage

How many natives from i are willing to work in country j?

- Utility of staying in country i: $U_i(h) = w_i h + \varepsilon_i$ (where is a iid extreme-value distributed error term)
- Utility of migrating to country j: $U_j(h) = w_j h - C_{i,j}(h) + \varepsilon_j$
- Probability that an individual from country i wants to emigrate to j:

$$\Pr\left[U_j(h) = \max_k U_k(h)\right] = \frac{\widehat{N}_{i,j}}{N_i} = \frac{\exp\left[w_j h - C_{i,j}(h)\right]}{\sum_k \exp\left[w_k h - C_{i,k}(h)\right]}$$

- Desired ratio of emigrants/residents:

$$\frac{\widehat{N}_{i,j}}{\widehat{N}_{i,i}} = \frac{\exp\left[w_j h - C_{i,j}(h)\right]}{\exp\left[w_i h\right]}$$

where a hat stands for desired migration.

Diasporas: predictions of self selection (1) - size

Without policy restrictions, we would have

$$\ln\left[\widehat{N}_{i,j}(h)\right] = (w_j - w_i)h - c(d_{i,j}, h, M_{i,j}, x_i) + \ln\left[\widehat{N}_{i,i}(h)\right]$$

The number of individuals who want to emigrate to j:

- increases with the differential in skill prices, $(w_j - w_i)$
- increases with the diaspora size, $M_{i,j}$
- increases with the resident population of skill h, $\widehat{N}_{i,i}(h)$ (size effect)
- decreases with distances (geographical, cultural), $d_{i,j}$

Diasporas: predictions of self-selection (2) - selection bias

- Predicting a continuous number of emigrants, our model is an approximation of the "discrete-number" real world with $\widehat{N}_{i,j}(h) \in \mathbb{N}$
- If $\ln\left[\widehat{N}_{i,j}(h)\right] \prec 0$, less than one migrant wants to leave her country. Assume it means that the bilateral migration flow is nil. The probability that $\widehat{N}_{i,j}(h) = 0$ is

$$\Pr\left[(w_j - w_i)h - c(d_{i,j}, h, M_{i,j}, x_i) + \ln\left[\widehat{N}_{i,i}(h)\right] \prec 0\right]$$

- Possibility of a selection bias in the data. A two-stage procedure can be useful to solve this problem

Diasporas: predictions of self-selection (2) - selection bias

- Ideally, we are looking for an instrument affecting this probability but not the size of migration flows. Suppose a variable x_i such that $c_x^{'} \simeq 0$ when migration flows are positive. However, when x_i is below a given threshold, migration costs become too high and impede migration. Such a variable would be a good instrument in the first-stage.
- Difficult to find such a variable. We will use the existence/size of a diplomatic representation of country j at origin (affecting the cost of obtaining a passport).
- Alternative methods: Heckman without instrument (identification through the analytical specification) or Poisson regressions

Diasporas: Predictions of self-selection (3) - quality

- We also have

$$\frac{\partial \ln \left[\widehat{N}_{i,j}(h) / \widehat{N}_{i,i}(h) \right]}{\partial h} = w_j - w_i - c_h^{'} \succ 0$$

- Positive selection arises as wage premia increases and migration costs decrease with h.
- It does not mean that there are more skilled emigrants than unskilled emigrants. It means the the skilled have a higher propensity to migrate to rich countries. If the proportion of skilled amon natives is low, there will be more unskilled than skilled migrants abroad.

Diasporas: Predictions of self-selection (4) - diasporas

What can we say about diaspora externalities?

- First, existing diasporas boost current migration:

$$\frac{\partial \ln \left[\widehat{N}_{i,j}(h) / \widehat{N}_{i,i}(h) \right]}{\partial M_{i,j}} = -c_M^{'} \succ 0$$

- Second, existing diaspora ambiguously affect the 'positive selection' of migrants

$$\frac{\partial^2 \ln \left[\widehat{N}_{i,j}(h) / \widehat{N}_{i,i}(h) \right]}{\partial h . dM_{i,j}} = -c_{hM}^{''} \gtrless 0$$

- Diasporas reduce 'positive selection' if diaspora externalities are stronger for low-skill workers: $-c_{hM}^{''} \prec 0$.

Diasporas: out-selection stage

- Suppose host countries are characterized by an immigration policy based on economic migration and family reunion.
- An individual of human capital h has a probability $p_j h$ to leave as an economic migrants (p_j measure the selectivity of the economic migration program)
- If not accepted as an economic migrant (with probability $1 - p_j h$), each individual has a probability to migrate through family reunion program. This probability depends on the size of the diaspora abroad, as percent of the native population.
- The probability to be accepted in country j is denoted by

$$\pi_{i,j}(h) = p_j h + (1 - p_j h) q_j \left(\frac{M_{i,j}}{N_i} \right)^{\mu}$$

where q_j measures the generosity of family reunion programs, $\frac{M_{i,j}}{N_i}$ affects the probability that a young has a relative abroad and $\mu \leq 1$ is a parameter.

Diasporas: predictions of the out-selection stage

We have:

$$\frac{\partial \ln\left[\pi_{i,j}(h)\right]}{\partial p_j} = \frac{h\left[1 - q_j\left(\frac{M_{i,j}}{N_i}\right)^{\mu}\right]}{\pi_{i,j}(h)} \succ 0$$

$$\frac{\partial \ln\left[\pi_{i,j}(h)\right]}{\partial q_j} = \frac{(1 - p_j h)\left(\frac{M_{i,j}}{N_i}\right)^{\mu}}{\pi_{i,j}(h)} \succ 0$$

$$\frac{\partial \ln\left[\pi_{i,j}(h)\right]}{\partial h} = \frac{p_j\left[1 - q_j\left(\frac{M_{i,j}}{N_i}\right)^{\mu}\right]}{\pi_{i,j}(h)} \succ 0 \quad (!!!)$$

$$\frac{\partial \ln\left[\pi_{i,j}(h)\right]}{\partial\left[M_{i,j}/N_i\right]} = \frac{(1 - p_j h) q_j \mu\left(\frac{M_{i,j}}{N_i}\right)^{\mu-1}}{\pi_{i,j}(h)} \succ 0 \quad (!!!)$$

Diasporas: predictions of the out-selection stage

- The first derivative implies that the existence of skill-related economic migration programs reinforces the self-selection mechanism: migrants are positively selected
- The last derivative implies that diasporas increase the probability to be accepted abroad.
- Furthermore, we have

$$\frac{\partial^2 \ln\left[\pi_{i,j}(h)\right]}{\partial h.\partial\left[M_{i,j}/N_i\right]} = \frac{-p_j q_j \mu \left(\frac{M_{i,j}}{N_i}\right)^{\mu-1}}{\left[\pi_{i,j}(h)\right]^2} \prec 0$$

$$\frac{\partial^2 \ln\left[\pi_{i,j}(h)\right]}{\partial h.\partial q_j} = \frac{-p_j \left(\frac{M_{i,j}}{N_i}\right)^{\mu}}{\left[\pi_{i,j}(h)\right]^2} \prec 0$$

- Diasporas and family reunion programs reduce the positive selection of migrants

Diasporas: full selection model

- Aggregating self-selection mechanims and policy restrictions, we can predict the effective bilateral flow from i to j. Starting from $N_{i,j}(h) = \widehat{N}_{i,j}(h).\pi_{i,j}(h)$, we have:

$$\ln\left[N_{i,j}(h)\right] = (w_j - w_i)h - c(d_{i,j}, h, M_{i,j}, x_i) + \ln\left[\widehat{N}_{i,i}(h)\right] + \ln\left[\pi_{i,j}(h)\right]$$

- $\widehat{N}_{i,i}(h)$ is an unobservable variable. We proxy it using the observed stock of residents $N_{i,i}(h)$.and consider it as exogenous.
- In principle, we should solve a model with

$$N_{i,i}(h) = \widehat{N}_{i,i}(h) + \sum_j \left[1 - \pi_{i,j}(h)\right] \widehat{N}_{i,j}$$

Diasporas: full selection model

- Effect of diasporas on the intensity of migration

$$\frac{\partial \ln \left[N_{i,j}(h)\right]}{\partial M_{i,j}} = -c_M^{'} + \frac{(1-p_j)q_j\mu\left(\frac{M_{i,j}}{N_i}\right)^{\mu-1}\frac{1}{N_i}}{\pi_{i,j}(h)} \succ 0$$

- Effect of diasporas on the selection of migrants

$$\frac{\partial^2 \ln \left[\widehat{N}_{i,j}(h)\right]}{\partial h.\partial M_{i,j}} = -c_{hM}^{''} - \frac{p_j q_j \mu\left(\frac{M_{i,j}}{N_i}\right)^{\mu-1}\frac{1}{N_i}}{\left[\pi_{i,j}(h)\right]^2} \gtrless 0$$

Diasporas: full selection model

What are the implications for host countries?

- $\frac{M_{i,j}}{N_i}$ can be very small. At the individual level, the probability to have a relative abroad is small for many countries
- However, on the aggregate, the role of diasporas is important for host countries. The flow of immigrants of type h in country j is given by:

$$I_j(h) = \sum_j \widehat{N}_{i,j}(h).\pi_{i,j}(h)$$

Diasporas: full selection model

What are the implications for host countries?

- The selection of migrants is governed by

$$\frac{\partial I_j(h)}{\partial h} = \sum_j \left[\frac{\partial \widehat{N}_{i,j}}{\partial h} \pi_{i,j}(h) + \widehat{N}_{i,j}(h) p_j \left[1 - q_j \left(\frac{M_{i,j}}{N_i} \right)^{\mu} \right] \right]$$

- Accepting more skilled migrants does not necessarily have a strong impact on the structure of immigration. The impact is low if family reunion programs are generous and diasporas are large:

$$\frac{\partial^2 I_j(h)}{\partial h . \partial p_j} = \sum_j \left[\frac{\partial \widehat{N}_{i,j}}{\partial h} h + \widehat{N}_{i,j}(h) \right] \left[1 - q_j \left(\frac{M_{i,j}}{N_i} \right)^{\mu} \right]$$

EMPIRICAL RSULTS

Diasporas: Estimates for size

- Empirical specification for migration flows (proxied by the change in stocks):

$$ln\left[N_{i,j}(h)\right] = \alpha + \alpha_1 M_{ij} + \alpha_2 d_{i,j} + \alpha_3 w_j + \alpha_4 A_j + \gamma_i + \epsilon_{ij}$$

  where w_j and A_j captures the attractiveness of destination j
- Or model with fixed effects:

$$ln\left[N_{i,j}(h)\right] = \alpha + \alpha_1 \ln(M_{ij}) + \alpha_2 d_{i,j} + \gamma_j + \gamma_i + \epsilon_{ij}$$

- For the sake of comparison, specification for migration stock:

$$ln\left[M_{i,j}(h)\right] = \alpha + \alpha_1 d_{i,j} + \gamma_j + \gamma_i + \epsilon_{ij}$$

Diasporas: Estimates for size

Econometric problem: 31 percent of zero values in bilateral stocks (36 percent in bilateral flows)
OLS leads to inconsistent estimates
Alternative 1: use Poisson regression for the size equation
Alternative 2: add a selection equation - Heckman regression with an instrument (diplomatic representation at origin) or without instruments (risk of collinearity between Mills ratio and explanatory variables)

Diasporas: Estimates for size

Table 2. Explaining diasporas size and change in size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Flows	Flows	Flows	Flows	Flows, low skilled	Flows,low skilled	Stocks	Stocks,low skilled
Lagged diasp	0.620	0.616	0.699	0.831	0.778	1.192		
	(34.35)***	(26.60)***	(43.91)***	(23.44)***	(22.25)***	(6.90)***		
Col links	0.331	0.278	0.127	-0.051	0.153	-1.699	2.138	2.336
	(2.45)**	(2.14)**	(1.10)	(0.29)	(0.64)	(2.05)**	(17.86)***	(17.34)***
linguage	0.388	1.026	0.496	1.056	0.322	1.413	1.259	1.049
	(5.20)***	(10.02)***	(6.48)***	(8.34)***	(2.18)**	(3.23)***	(14.95)***	(10.94)***
Log(dist)	-0.408	-0.139	-0.448	-0.095	-0.613	0.057	-1.035	-1.133
	(9.04)***	(2.48)**	(10.69)***	(1.63)	(7.40)***	(0.31)	(24.66)***	(23.70)***
Shengen	0.168	0.065	0.277	0.599	-0.081	1.154	-0.157	-0.464
	(1.19)	(0.33)	(2.02)**	(2.56)**	(0.28)	(1.31)	(1.09)	(2.82)***
Immig. pol		0.035		0.035		0.015		
		(7.85)***		(6.71)***		(0.87)		
Social exp		-0.290		0.175		2.411		
		(2.25)**		(1.28)		(3.22)***		
Pop at dest		0.321		0.109		-0.131		
		(9.66)***		(2.30)**		(0.83)		
Wages at dest		0.028		0.040		-0.020		
		(3.70)***		(4.51)***		(0.75)		
Constant	3.750	-4.954	2.365	-6.119	1.388	-17.084	8.847	8.422
	(6.92)***	(3.96)***	(4.02)***	(5.07)***	(1.20)	(2.99)***	(13.36)***	(9.51)***
Observations	3608	3091	5760	4992	5760	4992	5760	5760
Dest dum	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Orig dum	yes	No	yes	No	yes	No	yes	yes
Method	OLS	OLS	Heckman	Heckman	Heckman	Heckman	Heckman	Heckman
Mills ratio	-	-	1.19	1.92	2.09	5.44	1.07	1.11
			(9.35)***	(7.65)***	(6.70)***	(3.76)***	(9.77)***	(8.75)***

Diasporas: Estimates for selection

- Measure of selection: $S_{ij} = \frac{M_{ij}(s)}{M_{ij}(u)}$
- Empirical specifications for selection:

$$\begin{aligned} ln(S_{ij}) &= \alpha + \alpha_1 \ln(M_{ij}) + \alpha_2 d_{i,j} + \alpha_3 w_j + \alpha_4 A_j + \gamma_i + \epsilon_{ij} \\ ln(S_{ij}) &= \alpha + \alpha_1 \ln(M_{ij}) + \alpha_2 d_{i,j} + \gamma_j + \gamma_i + \epsilon_{ij} \end{aligned}$$

- Empirical specifications for changes in selection:

$$\begin{aligned} \Delta \ln(S_{ij}) &= \alpha + \alpha_1 \ln(M_{ij}) + \alpha_2 d_{i,j} + \alpha_3 w_j + \alpha_4 A_j + \gamma_i + \epsilon_{ij} \\ \Delta ln(S_{ij}) &= \alpha + \alpha_1 \ln(M_{ij}) + \alpha_2 d_{i,j} + \gamma_j + \gamma_i + \epsilon_{ij} \end{aligned}$$

Diasporas: Estimates for selection

Table 3. Impact of diaspora on selection (ratio Skilled/unskilled)

	(1) Skill ratio	(2) Skill ratio	(3) Skill ratio	(4) Skill ratio	(5) Change SR	(6) Change SR
Lagged diasp	-0.171	-0.088	-0.194	-0.132	-0.143	-0.108
	(16.19)***	(8.47)***	(20.62)***	(11.83)***	(17.62)***	(11.47)***
Col. links	-0.042	-0.439	-0.022	-0.410	0.101	0.096
	(0.62)	(6.08)***	(0.32)	(5.21)***	(1.67)*	(1.46)
linguage	0.466	0.703	0.460	0.721	0.176	0.257
	(9.38)***	(11.03)***	(9.37)***	(11.68)***	(4.17)***	(4.95)***
Log(dist)	0.096	0.273	0.090	0.263	0.086	0.116
	(3.35)***	(10.17)***	(3.40)***	(9.96)***	(3.78)***	(5.25)***
Shengen	0.502	0.305	0.519	0.303	0.390	0.117
	(5.65)***	(3.14)***	(6.26)***	(2.97)***	(5.48)***	(1.37)
Immig pol		-0.014		-0.015		0.001
		(4.98)***		(5.52)***		(0.30)
Soc exp		-1.206		-1.253		-0.756
		(16.11)***		(20.12)***		(14.42)***
Pop. at dest		0.061		0.082		0.056
		(3.45)***		(4.58)***		(3.75)***
Wage at dest.		0.044		0.045		0.035
		(9.86)***		(10.47)***		(9.78)***
Constant	-1.109	0.002	-0.734	0.257	-1.250	-0.563
	(1.16)	(0.00)	(1.32)	(0.34)	(2.54)**	(0.87)
Dest dum	Yes	No	Yes	No	Yes	No
Orig dum	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	OLS	Heckman	Heckman	Heckman	Heckman
Mills			-0.380	-0.446	-0.10	-0.99
			(6.86)***	(7.37)***	(0.22)	(1.88)*
Obs	3604	3084	5760	4992	5760	4992
R-squared	0.60	0.45				

Diasporas: Estimates for concentration

- We focus on relative concentration (skilled - unskilled)
- Bilateral measure of relative concentration = individual component of the Herfindhal index:

$$RC_{i,j} = C_{ij}^{s} - C_{ij}^{u} = \left[\frac{M_{ij}^{s}}{\sum_i M_{ij}^{s}}\right]^2 - \left[\frac{M_{ij}^{u}}{\sum_i M_{ij}^{u}}\right]^2$$

- Empirical specifiations (example with fixed effects):

$$\begin{aligned} RC_{ij} &= \alpha + \alpha_1 \ln(M_{ij}) + \alpha_2 d_{i,j} + \gamma_j + \gamma_i + \epsilon_{ij} \\ \Delta RC_{ij} &= \alpha + \alpha_1 \ln(M_{ij}) + \alpha_2 d_{i,j} + \gamma_j + \gamma_i + \epsilon_{ij} \end{aligned}$$

Diasporas: Estimates for concentration

Table 4 : Explaining relative concentration

	(1) Rel conc	(2) Rel conc	(3) Rel conc	(4) Rel conc	(5) Change RC	(6) Change RC
Lagged diasp	-0.502	-0.294	-0.514	-0.347	-0.008	-0.008
	(5.87)***	(3.54)***	(9.67)***	(5.73)***	(16.05)***	(15.45)***
Col. links	-4.635	-7.085	-4.619	-7.008	-0.040	-0.043
	(4.68)***	(6.41)***	(10.69)***	(14.75)***	(9.93)***	(10.45)***
Linguage	0.338	0.373	0.321	0.369	-0.004	-0.005
	(0.84)	(0.78)	(1.09)	(1.02)	(1.58)	(1.75)*
Log(dist)	0.266	0.628	0.269	0.615	0.006	0.006
	(1.24)	(3.73)***	(1.69)*	(3.91)***	(3.78)***	(4.26)***
Shengen	-0.193	-0.076	-0.180	-0.068	0.002	0.001
	(0.50)	(0.16)	(0.36)	(0.11)	(0.49)	(0.26)
Pop. at dest		0.956		0.988		0.003
		(7.13)***		(9.33)***		(3.50)***
Immig pol		-0.014		-0.013		0.000
		(1.31)		(0.84)		(1.51)
Soc exp		-1.509		-1.573		0.002
		(4.38)***		(4.44)***		(0.52)
Wage at dest		0.217		0.217		0.001
		(7.69)***		(8.57)***		(4.68)***
Constant	5.607	-18.397	-3.240	-10.824	-0.037	-0.111
	(0.29)	(4.70)***	(1.19)	(2.77)***	(1.60)	(3.33)***
Dest dum	Yes	No	Yes	No	Yes	No
Orig dum	Yes	Yes	Yes	Yes	Yes	Yes
Method	OLS	OLS	Heckman	Heckman	Heckman	Heckman
Mills			-0.405	-0.680	-0.873	-1.684
			(1.07)	(1.94)**	(2.44)**	(6.12)***
Observations	3920	3367	5730	4966	5730	4966
R-squared	0.29	0.17				

Diasporas: Conclusion

- OLS (lower bound of the impact of diaspora due to selection bias): R^2=89%; 71% of the observed variability of the migration flows is explained by the diaspora effects.
- Diaspora effects explain 47% and 78% of the total and explained variability of the selection ratio in 2000.
- Since economic variables at destination also play a crucial role (see Grogger and Hanson), immigration policies have a limited impact on the size and structure of migration flows.