LEM
WORKING PAPER SERIES

Revisiting the Role of Social Networks as Determinants of International-Migration Flows: A Note

Giorgio Fagiolo *
Gianluca Santoni *

* Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy

2014/20   October 2014

Revisiting the Role of Social Networks as Determinants of International-Migration Flows: A Note[*]

Giorgio Fagiolo[†]   Gianluca Santoni[‡]

October 2014

Abstract

This note revisits the role of migrant social networks as determinants of bilateral-migration flows. We do so using a new database that covers about 190 world countries and features more accurate estimates of bilateral flows than those employed so far. Our battery of gravity-model exercises show that the impact of social networks is consistent and significant over different specifications, and in line with previous estimates. Furthermore, in presence of migrant networks at destination, geographical distance counts in explaining the absence of a migration corridor only when such networks have very small sizes.

Keywords: International Migration; Gravity Models; Stocks vs Flows; Migrant Social Networks.

JEL Classification: F22; O15.

1   Introduction

International migration is increasingly shaping the world where we live by affecting many intertwined spheres of the demographic, social and economic fabric of both origin and destination countries (IOM, 2010). Quantifying the determinants of bilateral flows of migrants is therefore crucial to provide policy makers with the right tools (UNDP, 2009). Traditionally, this task has been performed fitting gravity models to longitudinal data and identifying the importance of population size, geographical distance, and other socio-political and economic factors potentially affecting the way in which prospective migrants choose their destinations (see Anderson, 2010; Beine et al., 2014, for an introduction). Among these potential determinants, quite a relevant role is played by social networks of migrants, whose presence at destination (Massey et al., 1998) lowers migrant expected risk and transaction costs associated with relocation.

One of the major challenges in identifying international-migration determinants is the absence of a longitudinal dataset of migration *flows* covering a large percentage of world countries. Whereas many studies have used migration-flow data recording immigration and emigration from and to a small subset of countries (Lewer and Van den Berg, 2008; Mayda, 2010; Kim and Cohen, 2010; Fitzgerald et al., 2014), others have employed a proxy of flows based on variations of migration stocks (Beine et al., 2011; Bertoli and Moraga, 2013; Beine and Parsons, 2012), which account for the number of foreign-born people living at a certain date in a given country. As discussed in Beine et al. (2014), this may introduce a severe bias, let alone choosing how to deal with negative stock variations. Other works have used stocks instead of flows (or estimates thereof) as dependent variable in the gravity model (Letouzé et al., 2009; Docquier et al., 2010; Felbermayr et al., 2010; Belot and Hatton, 2012; Ortega and Peri, 2013; Ramos and Surinach, 2013; Figueiredo et al., 2014). Although migration stocks can be interpreted as the representation of a long-term equilibrium and are probably of a better quality than flow data (Brücker and Siliverstovs, 2006), this choice is inconsistent with a proper microfoundation of the gravity model.

In this note, we revisit the issue of identifying the determinants of international-migration flows in general, and the role of social networks in particular, using a recent dataset of bilateral international-migration flows (Abel, 2013; Abel and Sander, 2014). This new database partly

overcomes the limitations of existing ones, in that it employs more sophisticated and consistent techniques to reliably estimate flows from stocks, while at the same time allowing for an almost-total country coverage for the period 1960-2000. We fit a battery of gravity models to bilateral flows, controlling for migrant social networks (i.e., migrant stocks at destination) and properly dealing with potential sources of endogeneity. We are mostly interested in assessing the network effect on migration flows, but also the role of geographical distance, in explaining zero migration-flow observations, in presence of positive network effects.

2   Data and Methods

We employ migration-flow estimates from Abel (2013), who provides bilateral-flow matrices for $N = 191$ countries over 5 decades from 1960 to 2000. Estimates are based on World Bank global-migration stock tables (Özden et al., 2011).[1] Table 2 provides summary statistics for $(t - 10, t]$ flow data (labeled with t) and their correlation with migration stocks in year $t - 10$, where $t = 1970, \ldots, 2000$. As expected, the number of (positive) flows increases in time, and the percentage of zeros decreases. The flow distribution appears to be very right-skewed, as the median is of two orders of magnitude smaller than the mean. Note also that flows are positively but not strongly correlated with stocks, and such correlation becomes weaker as we go back in time.

Our baseline gravity panel regression reads:

$$F_{ijt} = \beta_0 \cdot \prod_{h=1}^{H} Z_{iht}^{\beta_{1h}} \cdot d_{ij}^{\beta_2} \cdot S_{ij,t-10}^{\beta_3} \cdot \exp\left\{\mu_i + \eta_{jt} + \beta_4 \delta_t + \beta_5 \mathbf{X}_{ijt}\right\} \epsilon_{ijt}, \quad (1)$$

where F_{ijt} is the emigration flow from i to j in the time interval $(t - 10, t]$; Z_{iht} are push effects (e.g. population, per-capita GDP at origin, major episodes of political violence); d_{ij} is geographical distance; $S_{ij,t-10}$ are migration stocks controlling for social-network effects (i.e. number of people born in i and living in j at time $t - 10$); μ_i are fixed origin effects; η_{jt} are time-destination fixed effects controlling (also) for multilateral resistance (Bertoli and Moraga, 2013); δ_t are time

---

[1]Migration flows are estimated via a statistical procedure aimed at predicting the number of movements required to meet the changes over time in migrant stock data, using an iterative proportional fitting algorithm (Deming and Stephan, 1940).

dummies; $\mathbf{X}_{ijt}$ are (constant or time-varying) bilateral covariates such as geographical contiguity, common language, and colonial links; and ϵ_{ijt} is a shock with $E(\epsilon_{ijt}|\cdot) = 1$. Table 1 lists all our main variables and data sources.

The baseline specification above focuses on the identification of push effects, controlling for multilateral resistance using time-destination fixed effects (FEs). We also experiment with alternative FE combinations: (i) time-invariant FE only (μ_i, μ_j); (ii) time-varying origin and destination FEs (η_{it}, η_{jt}); (iii) time-varying origin FE and time-invariant pull effects (μ_j, η_{it}).

Furthermore, as many microfounded gravity models derive expressions in terms of migration (flow) rates (Beine et al., 2014), we replace in Eq. (1) the dependent variable with $\tilde{F}_{ijt} = F_{ijt}/P_{it}$ and migrant stocks with $\tilde{S}_{ijt} = S_{ij,t-10}/P_{j,t-10}$. This allows us to check whether migrant networks have an impact on bilateral-migration rates, net of any size effect.

Estimation is carried out using standard OLS (after log-linearization of Eq. 1), Poisson pseudo maximum-likelihood (Silva and Tenreyro, 2006), and a censored quantile-regression procedure (Figueiredo et al., 2014).

We control for potential endogeneity of migration stocks and income per capita using an IV approach. Endogeneity here can come both from an omitted-variable bias and, in the case of migration stocks, also from the fact that flows are computed using stocks. Although reverse causation should not be a concern for stocks, as the latter are predetermined with respect to flows, migration can feed back on income per capita, especially in poor and agriculturally-dependent countries. We check the robustness of our results in two ways. First, we instrument per-capita income with climatic anomalies, i.e. rain shortages and excess temperature (in addition to second-stage controls featured in Eq. 1). Second, to deal with endogeneity of migration stocks, we focus on the 1980-2000 period, and run an additional set of IV regressions that employ an internal instrument *à la* Card (Altonji and Card, 1991), i.e. stocks in 1960, plus climatic anomalies.[2] In both cases, Sargan-Hansen over-identification tests cannot reject the null hypothesis that all our instruments are valid.

A further issue that we address is how migrant networks and distance (in addition to the other

---

[2]In line with Beine and Parsons (2012), we do not find any consistent and significant effect of climatic anomalies on migration flows, when they are inserted in the migration-flow equation. Instead, they are a valid and strong instrument for both per-capita income and migration stocks.

covariates) affect the probability of finding a zero flow. We do so by fitting logit models where the dependent variable is the probability of observing a zero flow in presence of a strictly positive stock. In this case, we restrict estimation to the correspondent sample of flows and stocks, and we employ the same specifications that we use for our full-fledged gravity specifications.

3   Results

Our main results are summarized in Table 3, where we report estimates for the coefficients (β_2, β_3) —and marginal effects of networks and distance for logit models— for different combinations of FE assumptions, estimation techniques, and sample restrictions on stocks and flows. Before commenting them, some remarks are in order. First, the number of observations ranges from 68,608 (log-linearized OLS) to 145,160 ($=4\cdot191\cdot190$, for PPML). These numbers may decrease a little due to covariate unavailability. Second, all regressions reach a high goodness of fit (R^2 for OLS between 0.486 and 0.711). Third, all covariates consistently have the expected sign.[3] In particular emigration flows from i to j increase the larger origin population size; if the origin experienced a major episode of violence; and whether i and j share a border, a common language or colonial past. Income at destination positively affect migration flows, whereas the effect of income at origin is positive but weak, indicating that credit constraints may be at work.[4]

Coming back to Table 3, we note that the effect of migrant networks on bilateral flows in the unrestricted sample (columns 4-8) appears to be robust and consistent throughout all our exercises. A 10% increase in the stock of migrant at destination implies on average a 4% increase in bilateral migration flows. This confirms previous results in Beine et al. (2011); Beine and Parsons (2012); Bertoli and Moraga (2013). A persistently slightly higher figure (between 4.5% and 5%) is implied by instrumenting migration stocks using 1960 stock observations and climatic anomalies (IV2) and by CQR (5%). In the latter case, we find that the impact on the 3rd flow quartile is much higher ($\sim 6.5\%$), indicating that the effect of networks is non-monotonic and increases with the

---

[3]We have also tried to use additional covariates (e.g. unemployment, relative quality of institutions (Marshall and Jaggers, 2010), etc.), but they either substantially reduced our sample size, or turned out to be consistently not significant.

[4]This result is robust to inclusion of higher-order income terms (Mayda, 2010). We did not account for poverty indicators as in Belot and Hatton (2012) because that would have substantially reduced the sample size.

magnitude of flows.[5] As to the effect of distance on flows, OLS estimates consistently predict an 8% flow decrease due to a 10% increase in origin-destination distance. This effect decreases to about 5% according to PPML and 6-7% with CQR.

When we try to explain the existence of zero flows in presence of network effects, the latter lose their significance (column 9). This means that the probability that a migration corridor exists is independent of migrant networks, as the latter are only able to reinforce an existing migration relationship. In other words, the network effect mainly operates over intensive rather than extensive dimensions.

The distance effect is instead positive and significant, suggesting that the more distant apart are the two countries, the larger the probability that the corridor is absent when the stock is present at destination.[6]

Despite migrant networks do not seem to directly affect the probability of a zero flow, an interesting relationship ties them together with distance. In Figure 1 we plot the effect of a unit change in $\log(d_{ij})$ on the probability of a zero migration flow at different percentiles of the $\log(M_{ij,t-10})$ distribution. The decreasing pattern suggests that geographical distance has a significant effect on $\text{Prob}(F_{ij,t} = 0)$ only when bilateral stocks are very small. Put it differently, in presence of migrant networks at destination, distance counts in explaining the absence of a migration corridor only when such networks have very small sizes.

Finally, the impact of migrant networks on flows is quite robust to size effects. Columns (10)-(12) show indeed that estimated coefficients for the relationship between bilateral-flow ratios (flow over population at origin) and stock ratios (bilateral stock over population at destination) are well in line with those computed for levels.

---

[5]A cautionary note in interpreting this result is needed, because the dependent variable is censored at the median.

[6]The probability of a zero flow given a positive stock also increases the smaller population and per-capita GDP at origin, if the origin did not experiece violence episodes, and if the two countries do not share a common language. Note also that the probability of a strictly-positive flow in absence of network effects (S=0) decreases, as expected, the farther away the two countries are.

4 Concluding Remarks

In this note we have exploited a new international-migration database to revisit the impact of migrant social networks on bilateral-migration flows. The database covers about 190 world countries and contains more accurate estimates of bilateral flows than those employed so far (Abel, 2013).

Our gravity exercises confirm that migrant social networks at destination consistently enhance migration flows. Furthermore, we find that this effect mostly has an intensive form. Extensively, the presence of migrant networks does not affect the probability that a positive migration flow is present. The extensive margin is instead affected by geographical distance, which however impacts on the probability that a corridor is absent only when the size of migrant networks are small.

The above results are robust to a number of alternative specifications. First, the coefficient of the log of bilateral stocks stays close to 0.4 also when we perform cross-section regressions. The only exception concerns the 1960 stock, which has a 0.2 elasticity when inserted as explanatory variable of both the 1990-2000 and the 1960-1970 flow. In general, the stock coefficient decreases the more distant in time is from the time-interval where flows are computed. Second, some weak non-linearity is present in the relation between migration stocks and flows, as the impact of the former on the latter is marginally decreasing as stocks get larger. However, the elasticity is always positive in the entire range of observed bilateral stocks. Finally, coefficients close to 0.4 are recovered also when flow ratios ($\tilde{F}$) are regressed against stock levels (S) and when flow levels (F) are regressed against stock ratios ($\tilde{S}$).

References

Abel, G. J. (2013), "Estimating global migration flow tables using place of birth data", *Demographic Research*, 28: 505–546.

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

Altonji, J. G. and D. Card (1991), "The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives", in *Immigration, Trade, and the Labor Market*NBER Chapters, National Bureau of Economic Research, Inc, 201–234.

Anderson, J. E. (2010), "The Gravity Model", Working Paper 16576, National Bureau of Economic Research.

Beine, M., S. Bertoli and J. Moraga (2014), "A practitioners' guide to gravity models of international migration", Working Papers 2014-03, FEDEA.

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

Beine, M. and C. Parsons (2012), "Climatic factors as determinants of International Migration", Discussion Papers (IRES - Institut de Recherches Economiques et Sociales) 2012002, Universitè catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES).

Belot, M. V. and T. J. Hatton (2012), "Immigrant Selection in the OECD", *Scandinavian Journal of Economics*, 114: 1105–1128.

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

Brücker, H. and B. Siliverstovs (2006), "On the estimation and forecasting of international migration: how relevant is heterogeneity across countries?", *Empirical Economics*, 31: 735–754.

Deming, W. E. and F. F. Stephan (1940), "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known", *The Annals of Mathematical Statistics*, 11: 427–444.

Docquier, F., C. Ozden and G. Peri (2010), "The Wage Effects of Immigration and Emigration", Working Paper 16646, National Bureau of Economic Research.

Felbermayr, G. J., S. Hiller and D. Sala (2010), "Does immigration boost per capita income?", *Economics Letters*, 107: 177 – 179.

Figueiredo, E., L. Lima and G. Orefice (2014), "Migration and Regional Trade Agreement: a (new) Gravity Estimation", Working Papers 2014-13, CEPII.

Fitzgerald, J., D. Leblang and J. C. Teets (2014), "Defying the Law of Gravity: The Political Economy of International Migration", *World Politics*, 66: 406–445.

IOM (2010), *World Migration Report 2010*, International Organization for Migration, Geneva.

Kim, K. and J. E. Cohen (2010), "Determinants of International Migration Flows to and from Industrialized Countries: A Panel Data Approach Beyond Gravity1", *International Migration Review*, 44: 899–932.

Letouzé, E., M. Purser, F. Rodriguez and M. Cummins (2009), "Revisiting the Migration-Development Nexus: A Gravity Model Approach", MPRA Paper 19227, University Library of Munich, Germany.

Lewer, J. J. and H. Van den Berg (2008), "A gravity model of immigration", *Economics Letters*, 99: 164–167.

Marshall, M., G. and K. Jaggers (2010), "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010.", .

Massey, D., J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino and J. Taylor (1998), *Worlds in motion: Understanding international migration at the end of the millennium*, Oxford: Clarendon Press.

Mayda, A. M. (2010), "International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows", *Journal of Population Economics*, 23: 1249–1274.

Ortega, F. and G. Peri (2013), "Migration, Trade and Income", IZA Discussion Papers 7325, Institute for the Study of Labor (IZA).

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

Ramos, R. and J. Surinach (2013), "A Gravity Model of Migration between ENC and EU", IZA Discussion Papers 7700, Institute for the Study of Labor (IZA).

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

UNDP (2009), *HDR 2009 - Overcoming barriers: Human mobility and development*, Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Table 1: Variable Definition and Data Sources.

Variable	Definition	Source
Flows (F_{ijt})	Number of emigrants from i to j in the decade prior to year t	Abel (2013)
Stocks ($S_{ij,t-10}$)	Number of people born in i and living in j at year $t-10$	World Bank GBMD[b]
Distance (d_{ij})	Geographical (great-circle) distance between i and j	CEPII[b]
Population (P)	Country population ($t-10$)	WDI[c]
Flow Rate ($\tilde{F}_{ijt}$)	F_{ijt}/P_{it}	Abel (2013) and WDI[c]
Stock Rate ($\tilde{S}_{ij,t-10}$)	$S_{ij,t-10}/P_{j,t-10}$	WB GBMD[b] and WDI[c]
GDPpc	Per-Capita GDP ($t-10$)	WDI[c]
Conflict	=1 if country i experienced a major episode of violence at $t-10$	Center for Systemic Peace[d]
Contiguity	=1 if i and j share a border	CEPII[b]
Colony	=1 if i and j ever shared a colonial relationship	CEPII[b]
Language	=1 if i and j share a common spoken language	CEPII[b]
Excess Temperature	Positive values of temperature anomalies (observed values, standardized using 1901-2000 average and standard deviation)	BADC[e]
Rain Shortage	Absolute values of negative precipitation anomalies (observed values, standardized using 1901-2000 average and standard deviation)	BADC[e]

[a]Global Bilateral Migration Database, see `http://data.worldbank.org/data-catalog/global-bilateral-migration-database`; [b]See `cepii.fr`; [c]World Development Indicators (`data.worldbank.org`); [d]See `systemicpeace.org`; [e]British Atmospheric Data Centre, see `badc.nerc.ac.uk`.

Table 2: Descriptive Statistics

	Flows			
	1970	1980	1990	2000
# Destinations*	187	186	188	180
# Origins*	188	184	185	189
# Flows	15210	16325	18238	18835
% of Zeros	0.58	0.55	0.50	0.48
Mean	2122	2577	2331	2873
Median	12	13	14	16
Std. Dev.	23714	48788	27774	42202
	Correlation with Stocks			
1960	0.49	0.36	0.26	0.15
1970	. . .	0.34	0.33	0.19
1980	. . .	. . .	0.53	0.37
1990	. . .	. . .	. . .	0.54

 Note: Balanced Panel (191 countries). Mean, median and standard deviation computed for strictly-positive flows. Correlations are significant at p<0.01. (*): No. of countries with at least one positive flow.

Table 3: The impact of migrant networks and geographical distance on bilateral international-migration flows: Estimates of regression coefficients.

Fixed Effects				F≥ 0, S≥ 0					F=0, S> 0	$\tilde{F} \geq 0, \tilde{S} \geq 0$		
				OLS (F>0)			PPML	CQR^a	Logit^b	OLS ($\tilde{F} > 0$)	PPML	CQR
i	j	it	jt	NO	IV1	IV2	NO	NO	NO	NO	NO	NO
(1)	(2)	(3)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
							Migrant Networks					
Y	Y	N	N	0.373	0.376	0.452	0.371	0.479	0.000^c	0.373	0.298	0.445
Y	N	N	Y	0.413	0.395	0.466	0.410	0.547	-0.007	0.406	0.334	0.347
N	Y	Y	N	0.395	0.412	0.470	0.404	0.525	-0.001^c	0.394	0.332	0.378
N	N	Y	Y	0.441	N/A^d	0.487	0.417	0.580	-0.009	0.432	0.355	0.399
							Geographical Distance					
Y	Y	N	N	-0.816	-0.811	-0.704	-0.555	-0.847	0.082	-0.816	-0.773	-0.782
Y	N	N	Y	-0.789	-0.810	-0.711	-0.497	-0.670	0.075	-0.811	-0.762	-0.659
N	Y	Y	N	-0.811	-0.789	-0.703	-0.551	-0.709	0.088	-0.819	-0.776	-0.685
N	N	Y	Y	-0.798	N/A^d	-0.719	-0.577	-0.608	0.077	-0.831	-0.873	-0.626

Notes: Dependent variable: bilateral migration flows. F=flows. S=stocks. $\tilde{F}$=Flows over population at origin. $\tilde{S}$= Stocks over population at destination. IV1: Only per-capita GDP is instrumented using climate anomalies and other covariates. IV2: instrumentation of migration stocks using 1960 stock observations and climatic anomalies (sample: 1980-2000). All regressions include a constant, not reported. PPML: Poisson pseudo maximum-likelihood. CQR: Censored quantile regression (^{a}computed at the median). ^{c}Not significant at p=0.10. All remaining coefficients are significant at p=0.01. Standard errors clustered by country pairs. ^{b}Marginal effect at the mean. ^{d}with both (it;jt) FEs income at origin and/or destination cannot be inserted.

Figure 1: Zero flows in presence of network effects. Marginal effect of a unit change in the log of geographical distance on the probability of observing zero flows as a function of bilateral-stock percentiles. *Note*: Origin time-invariant and destination time-varying fixed effects included. All other covariates at their mean values.