Linguistic Distance, Networks and the Regional Location Decisions of Migrants to the EU

Julia Bredtmann[1], Klaus Nowotny[2], and Sebastian Otten[3]
[1]RWI, IZA, CReAM; [2]University of Salzburg, WIFO; [3]UCL, CReAM, RWI

Abstract

This paper analyzes the interaction between migrant networks and linguistic distance in the location decisions of migrants to the European Union at the regional level. We test the hypothesis that language and networks are substitutes in the location decision. Based on individual level data and a random utility maximization framework we find that networks have a positive effect on location decisions while the effect of linguistic distance is, as expected, negative. We also find a positive interaction effect between the two variables: networks are more important the larger the linguistic distance between the home and host countries, and the negative effect of linguistic distance is smaller the larger the network size.

Introduction

Motivation

- Migrant networks (diasporas) are among the most important determinants of international migration flows (e.g., Beine, Docquier & Özden 2011)
- Provide positive externalities for members of the same ethnic group that reduce migration costs
- An emerging literature has identified language as another important factor for migrants' location decisions (e.g., Adsera & Pytlikova 2015)
- Acquiring a foreign language is easier if the native language is linguistically closer to the language to be learned (Isphording & Otten 2014, Chiswick & Miller 2015)

Hypothesis

Networks and linguistic distance may be interdependent:

- Importance of migrant networks increases with linguistic distance
- Negative effect of linguistic distance decreases with the size of migrant networks

Research question

What is the role of network size and linguistic distance in the regional location decision of migrants?

Contribution

I. Analyze the interaction between migrant networks and linguistic distance
II. Model the location decisions of immigrants to the EU at the regional level using a large set of sending countries
III. Using a linguistic distance matrix for sending country-receiving region dyads (at NUTS-2 level) to capture within-country variation in linguistic distance

Modeling Location Choice

Random utility model

- Migrant i from sending country s faces a set of alternative receiving regions K. Utility of region $r \in K$:

$$
\begin{aligned}
u_{isr} &= V_{isr} + \varepsilon_{isr} \\
&= \beta_1 Network_{sr} + \beta_2 LD_{sr} + \beta_3 Network_{sr} \times LD_{sr} + \gamma' X_{isr} + \varepsilon_{isr}
\end{aligned}
$$

- X_{isr}: control variables specific to i, s, r, and sr dyad
- Behavioral model: choose $r \in K$ if and only if $u_{isr} \geq u_{isk}\ \forall\ k \in K$
- Assuming $\varepsilon_{isr} \sim i.i.d.$ extreme value, probability that i chooses r can be estimated by a Conditional Logit model (McFadden 1974):

$$
\Pr(y_{isr} = 1) = \frac{\exp(V_{isr})}{\sum_{k=1}^{K} exp(V_{isk})}
$$

- Log-likelihood function largely similar to Poisson ➔ PPML estimation

Data

Special evaluation of 2007 European Union Labour Force Survey (EU-LFS)

- Data represent about 7.4 million migrants
- 156 sending countries, 200 receiving NUTS-2 regions ➔ 31,200 *sr* dyads

Linguistic distance

- Average phonetic similarity between most commonly spoken language in the sending country and in the receiving region (Isphording & Otten 2014)
- Based on the Levenshtein distance

Results

Table 1: PPML Estimation of Migration Flows to the EU

	I Coef/StdE	II Coef/StdE	III Coef/StdE	IV Coef/StdE	V Coef/StdE
Network	0.4033†	0.2758†	0.1300†	0.1151†	0.1114†
	(0.0237)	(0.0469)	(0.0295)	(0.0269)	(0.0265)
LD	−0.0238†	−0.0348†	−0.0334†	−0.0231†	−0.0191†
	(0.0033)	(0.0045)	(0.0029)	(0.0032)	(0.0033)
Network × LD	–	0.0016†	0.0019†	0.0015†	0.0014†
		(0.0005)	(0.0003)	(0.0003)	(0.0003)
ln(distance)	–	–	–	−0.3823***	−0.4592†
				(0.1263)	(0.1258)
Colony	–	–	–	0.9714†	0.4476†
				(0.1138)	(0.1112)
Common off. lang.	–	–	–	–	0.9558†
					(0.1267)
Constant	−4.0581†	−3.3798***	−5.3086†	−3.4647***	−3.8880***
	(1.0487)	(1.0542)	(1.0986)	(1.3054)	(1.3140)
Sending-country FE	yes	yes	yes	yes	yes
Receiving-region FE	no	no	yes	yes	yes
R^2	0.417	0.409	0.671	0.703	0.708
Observations	31,194	31,194	31,194	31,194	31,194

*Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – Robust standard errors in parentheses. – LD: linguistic distance. – PPML: Poisson pseudo-maximum-likelihood.*

- Interaction effect both statistically and economically significant

Heterogeneous effects

- If LD=0, a one standard deviation increase in Network increases the probability of migrating to that region by 19%.
- At the 90th-percentile of the LD distribution, a one standard deviation increase in Network increases the probability of migrating to that region by 50%.

Robustness Checks

Table 2: PPML Estimation of Migration Flows – Robustness Checks

	I Coef/StdE	II Coef/StdE	III Coef/StdE	IV Coef/StdE
Network	0.1109†	−0.1698	–	–
	(0.0265)	(0.1033)		
LD	−0.0192†	−0.0656†	−0.0143†	−0.0728†
	(0.0033)	(0.0071)	(0.0024)	(0.0099)
Network × LD	0.0014†	0.0045†	–	–
	(0.0003)	(0.0011)		
Relative network	–	–	0.0205***	–
			(0.0066)	
Relative network × LD	–	–	0.0006†	–
			(0.0001)	
Linguistic network	–	–	–	−0.1579
				(0.1220)
Linguistic network × LD	–	–	–	0.0043†
				(0.0012)
ln(distance)	−0.4587†	−0.4658†	−0.8535†	−0.4300†
	(0.1270)	(0.1293)	(0.1385)	(0.1281)
Colony	0.4438†	0.4631†	0.5711†	0.5061†
	(0.1113)	(0.1105)	(0.1075)	(0.1111)
Common off. lang. (COL)	0.9609†	1.0273†	1.1553†	1.0924†
	(0.1268)	(0.1362)	(0.1308)	(0.1288)
Genetic distance	−0.0210	–	–	–
	(0.2940)			
Constant	−1.5408	0.0888	−1.9284	−0.8117
	(1.2467)	(1.3551)	(1.3655)	(1.4916)
Sending-country FE	yes	yes	yes	yes
Receiving-region FE	yes	yes	yes	yes
Sample LD = 0 incl.	yes	no	yes	no
R^2	0.708	0.675	0.649	0.670
Observations	30,794	30,451	31,194	30,451

*Notes: – † $p < 0.001$; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. – Robust standard errors in parentheses. – LD: linguistic distance. – PPML: Poisson pseudo-maximum-likelihood. – Information on genetic distance is not available for Andorra an the State of Palestine, reducing the sample by 400 observations. – Relative network is calculated as the stock of migrants from sending country s living in region r divided by the total number of migrants from that sending country living in the EU, i.e., Relative network = $(stock_{sr}^{\leq 1998}/stockEU_{s}^{\leq 1998}) \times 100$. – Linguistic network is calculated as the stock of migrants living in region r that were born in a country that has the same most common language as the migrant's country of birth s.*

Multilateral resistance

- PPML requires error terms to be cross-sectionally independent
- Violated if V_{sr} fails to include all relevant bilateral determinants or if observed factors have heterogeneous impact across decision makers
- Multilateral resistance can be interpreted as a violation of independence of irrelevant alternatives (IIA)

Further robustness check

- Relax IIA and model source of heterogeneity
- Estimate Mixed Random Parameters Logit model

➔ Results confirm the findings of the PPML estimation

Conclusion

- Networks and linguistic proximity are substitutes in migrants' location choice
- Networks are more important when linguistic distance is high, and the negative effect of linguistic distance is smaller when networks are large

Contact

Sebastian Otten
University College London (UCL)
Department of Economics
Centre for Research and Analysis of Migration (CReAM)
Email: s.otten@ucl.ac.uk

References

1. Adserà, A. and Pytliková, M. (2015). The Role of Language in Shaping International Migration: Evidence from OECD Countries 1985-2006. *Economic Journal*, 125 (586), F49–F81.
2. Beine, M., Docquier, F. and Özden, A. (2011). Diasporas. *Journal of Development Economics*, 95 (1), 30–41.
3. Chiswick, B. R. and Miller, P. W. (2015). International Migration and the Economics of Language. In B. R. Chiswick and P. W. Miller (eds.), Handbook of the Economics of International Migration 1A, Oxford and Amsterdam: North-Holland, pp. 211–269.
4. Isphording, I. E. and Otten, S. (2014). Linguistic Distance and the Language Fluency of Immigrants. *Journal of Economic Behavior & Organization*, 105, 30–50.
5. McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choices. In P. Zarembka (ed.), Frontiers in Econometrics, New York: Academic Press, pp. 105–142.