WHY PEOPLE MOVE?
DETERMINANTS OF MIGRATION I

Mariola Pytliková

CERGE-EI and VŠB-Technical University Ostrava,

CReAM, IZA, CCP and CELSI

Info about lectures: https://home.cerge-ei.cz/pytlikova/LaborSpring16/
Office hours: by appointment

Contact:
Email: Mariola.Pytlikova@cerge-ei.cz
Mobile: 739211312

https://sites.google.com/site/pytlikovaweb/

Study Materials and Reading List

- Slides of the lectures (provided one day in advance or on the day of the class)
- All materials provided on: http://home.cerge-ei.cz/pytlikova/LaborSpring16/

Compulsory Readings:

- Borjas 6e, Chapter 8 labor mobility;
- Adserà, Alícia and Mariola Pytliková (2015): "The Role of Language in Shaping International Migration". ***Economic Journal,*** Vol. 125, Issue 586, pp. F49-F81. August 2015.

Other Relevant Literature:

- Pedersen, J. P., Pytlikova, M. and N. Smith (2008): "Selection and Network Effects - Migration Flows into OECD Countries 1990-2000". ***European Economic Review.*** Vol. 52 (7), pp. 1160-1186.
- Clark, Hatton and Williamson (2007): "Explaining U.S. Immigration, 1971–1998". ***The Review of Economics and Statistics.*** May 2007, Vol. 89, No. 2, Pages 359-373,
- Munshi, K. (2003), "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market", ***Quarterly Journal of Economics***, Vol. 118 (2), pp. 549-599.

WHY DO PEOPLE MIGRATE? Theory I

- ***ECONOMIC FACTORS:***
  - Wage differences (Hicks, 1932; Kuznetz and Rubin, 1954),
  - **Human capital model (Sjaastad,1962; Becker,1964): Move if net discounted future expected benefits>costs of migration** (assumed to be proportional to distance), later formalization of the model – *a starting point to most of the literature on migration determinants.*
  - Sjastaad's framework includes features of gravity model by viewing distance as a proxy for migration costs

WHY DO PEOPLE MIGRATE? Theory I

- ***ECONOMIC FACTORS:***
  - Income expectations conditioned on probability of being employed (Harris & Todaro, 1970; Hatton, 1995),
  - Family or households decision (Mincer,1978),
  - Relative deprivation approach (Stark, 1984),
  - “Welfare magnet” (Borjas, 1999), or “social tourism”, “social raids” (Kvist, 2004).

WHY DO PEOPLE MIGRATE? Theory II

- ***MIGRATION NETWORKS:***
  - **migration networks:** "…sets of interpersonal ties that connect migrants, former migrants, and non-migrants in origin and destination areas through ties of kinship, friendship, and shared community origin" (Massey, 1993)
  - **help to explain persistence in migration**
  - **"herd behavior" effect** (Bauer et al. 2002),
- ***NON-ECONOMIC FACTORS*:**
  - **war, love/marriage, taste for adventure**
  - **Language proximity**
- ***OTHER (UN)OBSERVABLE COUNTRY SPECIFIC FACTORS***

WHY DO PEOPLE NOT MIGRATE? Theory

- **Less than 3-4 percent of the world's population is living in a country other than they were born.**

?? WHY THERE IS NOT THAT MUCH MIGRATION ??

- ***BARRIERS TO MIGRATION:***
  - Immigration policies
  - Costs of migration (out-of-pocket exp., psychological costs)
  - Cultural distance
  - Language barriers
  - Skill transferability

The role of language in shaping international migration

Alícia Adserà
Princeton University,
IZA and CReAM

Mariola Pytliková
CERGE-EI and VŠB-Technical University Ostrava,
IZA, CReAM, CERGE, CELSI and CCP

Motivation

- Purpose of the paper: to study the role of language in explaining international migration flows from multiple angles:
  - linguistic proximity,
  - widely spoken languages,
  - linguistic enclaves,
  - language-based immigration policy requirements.

Motivation

Linguistic proximity and widely spoken languages

- Language plays a key role in the transfer of human capital to a foreign country - it helps the immigrant to be successful at the destination country's labor market
  - see e.g. Kossoudji (1988), Dustmann (1994), Dustman and van Soest (2002), Chiswick and Miller (2002, 2007), Dustmann and Fabbri, (2003), and Bleakley and Chin (2004) .
- => the ability to learn quickly the destination language and linguistic proximity between destinations and origins facilitates the transfer of human capital and reduces migration cost
- => **linguistic skills and linguistic proximity seem to play an important role in driving international migration flows.**

Motivation – Linguistic enclaves

- The composition and diversity of migrants already in destination affect the likelihood of finding previous migrants from same country and/or linguistic groups.
- Networks and *linguistic enclaves* (even if not from same country) may facilitate labor market entry to newcomers
  - i.e. migrants for all Central America moving to highly Mexican areas in the US.
  - Many immigrants whole lives working in a linguistic enclave (i.e. Boyd 2010 for the case of Canada).

Motivation – Previous evidence

- Previous evidence on determinants of migration flows mostly limited to a simple dummy for a common language
  - E.g.: Clark, Hatton and Williamson (2007), Pedersen, Pytlikova and Smith, (2008), Mayda (2010), Grogger and Hanson (2011), Beine, Docquier and Ozden (2011).

- Only two studies with more sophisticated measures:
  - Belot and Hatton (2012) use the number of nodes on the linguistic tree between two languages. Belot and Ederveen (2012) employ the linguistic proximity index by Dyen et al. (1992). Both only for *within OECD migration flows*.

Motivation – Contributions of this paper

This paper….

A) New dataset on migration flows & stocks to 30 OECD countries from all world countries as well as new linguistic proximity indices.

B) Explore different dimensions of language-migration link:

1. Multiple indices of Linguistic Proximity
2. Role of English as widely spoken language
3. Linguistic enclaves,
4. Language-based immigration policy requirements

Separate paper:

5. Linguistic diversity in origin and destination

Model –based on "human capital investment" theoretical framework

- We assume that an individual *k* decides whether to stay in his/her country of origin *i* or whether to migrate from country *i* to any potential destination *j*, where $j = 1, 2, .., J.$
- We assume that a potential immigrant maximizing her utility chooses to locate in the country where her utility is the highest among all available destinations.
- The utility that migrant *k*, currently living in *i*, attains by moving to *j* is logarithmic and given by:

$$U_{kij} = (y_{kj} - c_{kij})^{\lambda} \exp(\varepsilon_{kij}) \qquad (1)$$

- Where $y_{kj} - c_{kij}$ is the difference between income in destination *j*, (which can be defined in line with Harris and Todaro (1970) as wage times the probability of finding a job, *y* = *we*), and the cost of migrating from the home country *i* to *j*, c_{klj}

Model (based on Grogger-Hanson)

- We can write the probability of individual *k* from country *i* choosing a country *j* among *J* possible destinations as:

$$\Pr(j_k / i_k) = \Pr\left[U_{ijk} = \max(U_{ki1}, U_{ki2}, \ldots, U_{kiJ})\right] \qquad (2)$$

- Assuming that ε_{kij} follows an *i.i.d.* extreme value distribution and $\lambda>0$, and using the approximation that, $\ln(y_j - c_{ij}) \approx \ln y_j - (c_{ij} / y_j)$, we apply the results in McFadden (1974) to write the log odds of migrating to destination country *j* versus staying in the source country *i* as follows:

$$\ln \frac{M_{ij}}{P_i} = \ln m_{ij} \approx \lambda[\ln y_j - \ln y_i] - \lambda C_{ij} \qquad (3)$$

- where M_{ij} are flows of individuals from *i* to *j*; P_i are the stayers; m_{ij} is the emigration rate from *i* to *j* and C_{ij} are migration costs expressed as a proportion of destination income, $C_{ij}=(c_{ij}/y_{ij})$.

$$\ln(y_j - c_{ij}) \approx \ln y_j - (c_{ij} / y_j),$$

Model (based on Grogger-Hanson)

The probability of migration depends on the difference between income related to staying at home country *i* or migrating abroad *j* adjusted for costs of migration. Costs of moving to foreign country may be three fold: direct out-of-pocket costs of migrating and psychological costs of leaving own country, family and friends, and costs associated with a loss of skills due to imperfect skill transferability,

Suppose that income in a location can be defined in line with Harris and Todaro (1970) as wage times the probability of finding a job $y = we$

where *e* denotes employment rate, *w* real earnings. Then the migration rate in (3) can be expressed in terms of employment rates and wages

$$\ln \frac{M_{ij}}{P_i} = \ln m_{ij} \approx \lambda[\ln y_j - \ln y_i] - \lambda C_{ij} = \lambda[\ln w_{kj} + \ln e_{kj} - \ln w_{ki} - \ln e_{ki}] - \lambda C_{ij} \quad (4)$$

Empirical Model

We use the model to derive:

$$\ln(\boldsymbol{m}_{ijt}) = \gamma_1 + \gamma_2 \ln(\boldsymbol{gdp}_{jt-1}) + \gamma_3 \ln(\boldsymbol{gdp}_{it-1}) + \gamma_4 \ln(\boldsymbol{u}_{jt-1}) + \gamma_5 \ln(\boldsymbol{u}_{it-1}) + \gamma_6 \ln(\boldsymbol{pse}_{jt-1}) + \\ + \gamma_7 \ln(\boldsymbol{s}_{ijt-1}) + \gamma_8 \boldsymbol{L}_{ij} + \gamma_9 \boldsymbol{D}_{ij} + \gamma_{10} \boldsymbol{FH}_{it-1} + \gamma_{11} \boldsymbol{lr}_{jt-1} + \gamma_{12} \ln(\boldsymbol{p}_{ijt-1}) + \boldsymbol{\delta}_j + \boldsymbol{\delta}_i + \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_{ijt}$$

- **mijt -** emigration rate = gross migration flow per source country population,
- ***j*** destination country; *j* = 1,…, 30; ***i*** source country; *i* = 1,…,225;
- **Sijt-1** is stock of immigrants per source country population
- **Dij** is matrix of distance variables reflecting costs of moving
- **Pse welfare expenditure; FH freedom house political and civil rights**
- **U is unemployment; GDP is per capita; p is population ratios**
- **Lij** is a matrix of linguistic variables
- **A set of year dummies,** destination and source country fixed effects included
- **uijt** error term – clustered on the level of pair of countries

Data & models– Flows and stocks of migrants

- New dataset on Immigration flows and foreign population stock into 30 OECD countries from 223 countries.
- *Currently an update for 42 destinations and 1980-2012 period*
- Collected by writing to national statistical offices.
- Period: 1980 to 2010.
- Unbalanced panel.
- Improvement w.r.t. to other datasets – e.g. Docquier and Marfouk (2006), OECD (2011), WB (2011), UN (2011):
  - Both flows and stocks annually
  - Comprehensive in destinations, origins and time

Appendix Table A3: Inflows of foreign population: definitions and sources

Migration flows to:	Definition of "foreigner"	Source
Australia	Country of Birth	Permanent and long term arrivals, Government of Australia, DIMA, Dept. of Immigration and Multicultural Affairs http://www.immi.gov.au/media/statistics/index.htm
Austria	Citizenship	Population register, Statistik Austria (1997 to 2002), Wanderungsstatistik 1996-2001, Vienna
Belgium	Citizenship	Population register. Institut National de Statistique.
Canada	Country of Birth	Issues of permanent residence permit. Statistics Canada – Citizenship and Immigration Statistics. Flow is defined as a sum of foreign students, foreign workers and permanent residents.
Czech Rep.	Citizenship	Permanent residence permit and long-term visa, Population register, Czech Statistical Office
Denmark	Citizenship	Population register. Danmarks Statistics
Finland	Citizenship	Population register. Finish central statistical office
France	Citizenship	Statistics on long-term migration produced by the 'Institut national d'études démographiques (INED)' on the base on residence permit data (validity at least 1 year) transmitted by the Ministry of Interior.
Germany	Citizenship	Population register. Statistisches Bundesamt
Greece	Citizenship	Labour force survey. National Statistical Service of Greece 2006-2007 Eurostat
Hungary	Citizenship	Residence permits, National Hungary statistical office.
Iceland	Citizenship	Population register. Hagstofa Islands national statistical office.
Ireland	Country of Birth	Labour Force Survey. Central Statistical Office. Very aggregate, only few individual origins.
Italy	Citizenship	Residence Permits. ISTAT
Japan	Citizenship	Years 1988-2005: Permanent and long-term permits. Register of Foreigners, Ministry of Justice, Office of Immigration. Years 2006-2008: Permanent and long-term permits. OECD Source International Migration data
Korea	Citizenship	OECD Source International Migration data
Luxembourg	Citizenship	Population register, Statistical Office Luxembourg
Mexico	Citizenship	OECD Source International Migration data
Netherlands	Country of Birth	Population register, CBS
New Zealand	Last Permanent Residence	Permanent and Long-term ARRIVALS (Annual – Dec)

Appendix Table A1: Country-year coverage migration flows

Year/ Dest	AUS	AUT	BEL	CAN	CHE	CZE	DEU	DNK	ESP	FIN	FRA	GBR
2010	208	190		217	198	135	193	203	113	183		
2009	205	190	184	214	194	141	193	203	113	183		26
2008	204	190	182	214	194	143	194	203	113	183	120	21
2007	206	190	93	214	194	147	193	203	113	183	124	19
2006	206	190	96	214	194	142	193	202	108	183	120	34
2005	203	190	85	214	194	142	191	203	66	183	107	114
2004	203	190	71	214	194	146	191	203	57	183	107	109
2003	201	189	70	214	195	142	191	203	57	183	127	107
2002	198	189	70	214	194	141	191	203	57	183	128	99
2001	198	189	70	214	194	115	84	203	57	183	130	106
2000	200	189	70	214	180	110	83	203	59	183	129	111
1999	198	189	70	214	180	108	193	203	58	183	118	110
1998	193	189	70	214	180	122	193	203	59	183	117	116
1997	192	189	55	214	179	111	193	203	39	183	118	48
1996	195	189	55	214	176	114	193	203	58	183	118	52
1995	187		55	214	176	117	193	203	39	183	118	54
1994	186		55	214	179	106	193	203	39	183	118	27
1993	180		48	214	178	97	193	203	39	183		39
1992	182		48	214	174		189	203	45	183		45
1991	171		48	213	158		172	203	42	183		49
1990	168		48	213	156		44	203	42	183		38
1989	155		48	213	154		105	203	42	183		31
1988	150		25	213	159		105	203	42	183		38
1987	159		27	213	155		105	203		183		29
1986	153		27	213	154		105	203		183		33
1985	155		27	213	154		105	203		183		35
1984	154		27	213	151		105	203		183		
1983	166		27	213	152		105	203		183		
1982	161		27	213	154		105	203				
1981			27	213	154		105	203				
1980			27	213			105	203				
	AUS	AUT	BEL	CAN	CHE	CZE	DEU	DNK	ESP	FIN	FRA	GBR

Appendix Table A4: Stock of foreign population: definitions and sources

Foreign population stock in:	Definition of "foreigner" based on	Source:
Australia	Country of birth	Census of Population and Housing, Australian Bureau of Statistics
Austria	Country of birth	Statistics Austria, Population Census 2001 and Population Register 2001 to 2009. For census year 1981 and 1991 definition by citizenship
Belgium	Citizenship	Population register. Institut National de Statistique
Canada	Country of birth	Census of Canada, Statistics Canada.
Czech Rep.	Citizenship	Permanent residence permit and long-term visa, Population register, Czech Statistical Office and Directorate of Alien and Border Police
Denmark	Country of origin	Population register. Danmarks Statistics
Finland	Country of birth	Population register. Finish central statistical office
France	Country of birth	Census. Residence permit. Office des migrations internationals.
Germany	Citizenship	Population register. Statistisches Bundesamt
Greece	Citizenship	Labour force survey. National Statistical Service of Greece.
Hungary	Citizenship	National Hungary statistical office
Iceland	Country of birth	Population register. Hagstofa Islands
Ireland	Country of birth	Censuses, Statistical office, Ireland
Italy	Citizenship	Residence Permits. ISTAT

Appendix Table A2: Country-year coverage migration stocks

Year/Dest	AUS	AUT	BEL	CAN	CHE	CZE	DEU	DNK	ESP	FIN	FRA	GBR
2010		209			191	171	192	201		193		179
2009	209	209	185		194	172	190	201	112	191		171
2008	209	209	187		194	171	192	201	112	191	127	177
2007	209	209	178		194	168	193	200	112	191	128	174
2006	199	209	184	210	194	168	193	200	112	193	193	148
2005	209	209	182		194	166	139	201	112	193	204	97
2004	208	209	181		194	165	139	201	112	193		101
2003	208	209	181		194	163	138	201	112	193		100
2002	208	209	181		194	161	138	201	99	193		100
2001	190	207	181	190	194	163	138	201	99	193		97
2000	207	191	176		195	161	138	201	99	193		102
1999	206		174		195	164	138	201	99	193	162	87
1998	206		174		195	158	138	201	99	193		104
1997	204		55		195	152	138	201	99	193		100
1996	192		55	201	195	153	138	201	63	193		90
1995	202		55		195	150	138	201	58	193		85
1994	49		55		195	145	137	201	58	193		87
1993	49		48		195		137	201	58	193		87
1992	49		48		194		132	201	58	193		82
1991	168		48	180	194		117	201	58	193		70
1990	49	70	48		194		118	201	57	193	76	
1989			48		194		118	201	57	134		
1988					194		118	201	57	134		
1987					194		118	201	57	131		
1986	75			42	194		118	201	57	125		
1985					194		118	201	57	124		
1984					194		118	201		191		
1983					194		118	201				
1982					194		118	201				
1981	81		47	42	194		118	201				
1980		64			194		116	201				
Dest	AUS	AUT	BEL	CAN	CHE	CZE	DEU	DNK	ESP	FIN	FRA	GBR

Data & models– Flows and stocks of migrants

- Dependent variable: Ln Migration rates (flows normalized by population at origin *1000)
  - We add a one to immigration flows and foreign population stocks prior to constructing emigration and stock rates and taking logs, not discard the “zero” observations (only around 4.5 % in our data)

VARIABLES	Obs	Mean	Sd	Min	Max
Ln Emigration Rate	100519	-5.1221	2.5552	-14.0408	4.1193
Ln Stock of Migrants_t-1	102472	-3.1922	2.8966	-12.1770	6.5313

- Estimation: similar results across methods OLS pooled; random effects; OLS with year, origin and destination fixed effects (shown next).
- Poisson as robustness.

Controls in all models

- **Stock of immigrants per source country populations**
- **Distance variables reflecting costs of moving:**
  - Neighboring Country
  - Colonial past
  - Distance in Kilometers
  - Genetic distance (distance of distributions of alleles in both populations by Cavalli-Sforza, Menozzi, and Piazza 1994) - to rule out that language is masking other factors such as cultural or genetic similarity among populations.

Controls in all models

- **Socio-economic variables for receiving and sending countries:**
  - GDP per capita origin (& non-linear term to capture potential poverty traps) & destination,
  - Unemployment rates origin & destination
  - Public social expenditure in destination, %GDP in j,
  - Population ratio; receiving/sending,
  - Freedom House Indexes: political rights and civil liberties
- **Year, origin and destination fixed effects**

Building a Linguistic proximity variable Ethnologue –Linguistic Tree. Example from Desmet et al. (J. Development Ec 2012)

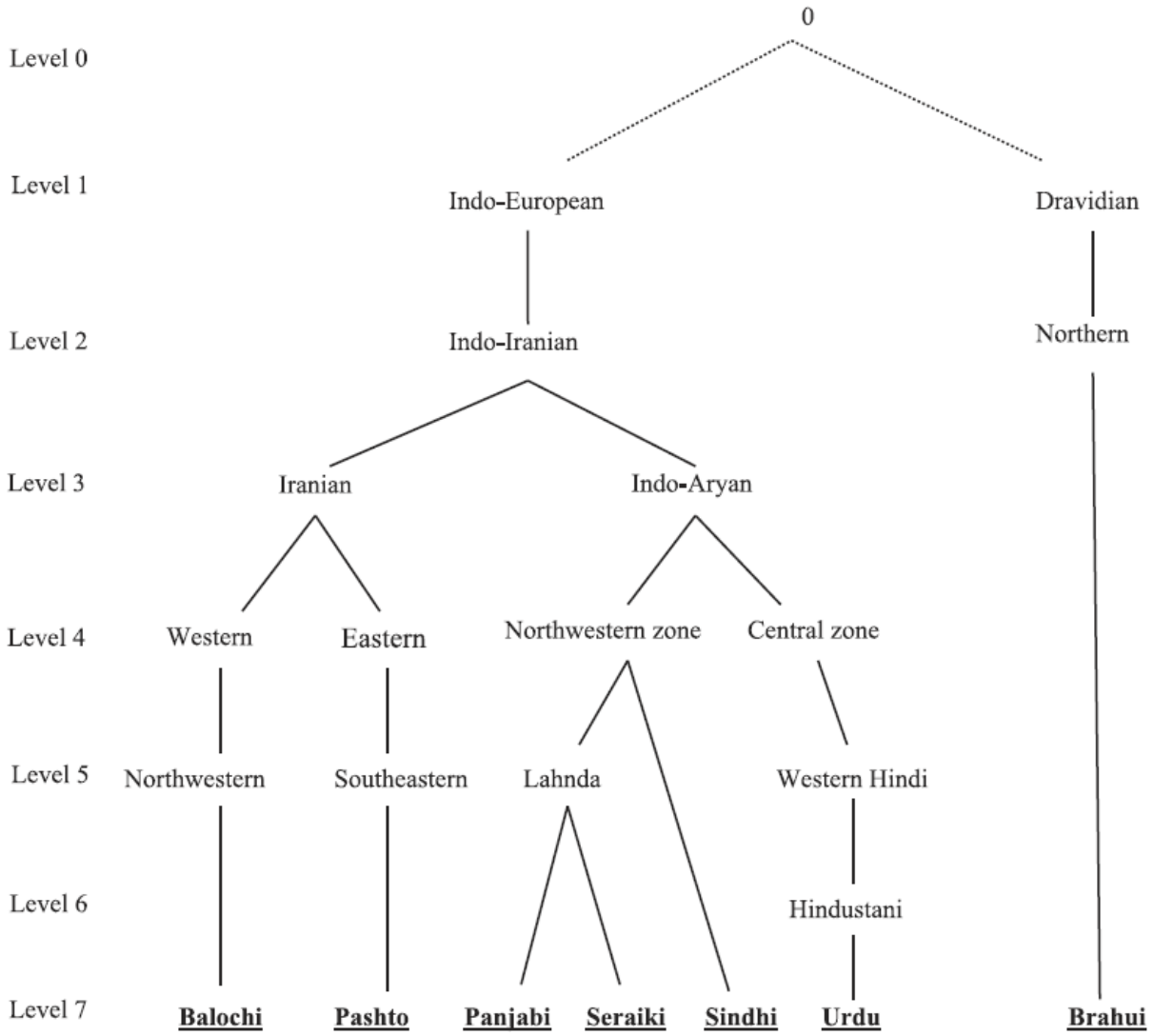


Fig. 1. Phylogenetic tree of major languages in Pakistan.

Building a Linguistic proximity variable

- ***Index ranges (0-1) depending on the highest level that two languages share in the family linguistic tree of Ethnologue Encyclopedia***
- 1) We define 4 weights up to the 4th level of the linguistic tree shared:
  - ***SAMEW1*= 0.1; *1st level:*** *e.g.* Indo-European versus Urallic (Fin, Est, Hun).
  - ***SAMEW2*= 0.15; *2nd level:*** e.g. Germanic versus Slavic
  - ***SAMEW3*= 0.20; *3rd level:*** e.g. Germanic W. vs. Germanic N.
  - ***SAMEW4*= 0.25; *4th level:*** e.g. Scandinavian W. (ISL) vs. Scandinavian E. or German vs. English.

- 2) Define the linguistic index by:
- **INDEX= *SAMEW1 + SAMEW2 + SAMEW3 + SAMEW4***

No Share=0; MaxShare1st=0.1; MaxShare2nd =0.25, MaxShare3rd =0.45; MaxShare4th =0.70; Same=1

Language proximity and ln. migration rates from 223 countries of origin to 30 OECD destination countries for 1980-2010.

Δ in St. Dev migration rates from Δ one St Dev 0.020*** (BETAS)

	OLS	OLS	FE	FE	Poisson
VARIABLES	(1)	(2)	(3)	(4)	(5)
Linguistic Proximity	3.271***	-	0.732***	0.209***	0.508***
	(0.147)		(0.123)	(0.066)	(0.127)
Common Language	-	2.929***	-	-	
		(0.169)			
Ln Stock of Migrants_t-1	NO	NO	NO	YES	YES
Economic controls	NO	NO	YES	YES	YES
Pop ration, Distance & political vars	NO	NO	YES	YES	YES
Destination & Origin FE	NO	NO	YES	YES	YES
Observations	100519	100519	74797	51257	51257
Adjusted R-squared	0.111	0.076	0.764	0.899	

↑R^2 with proximity index

Notes: Dependent Variable: Ln (Emigration Rate). Controls included: **stock of migrants, economic & political variables, distance variables, colonial, year dummies and destination and origin country fixed effects**. Robust standard errors clustered at the country-pair level, *** p<0.01, ** p<0.05, * p<0.1.

Interpretation 1980-2010

- Cols (4), our baseline spec: Emigration flows to a country with same language as opposed to one with no common family should be around 20% higher.
- When comparing emigration rates to France in (4):
- **Ceteris paribus, rates from Benin (with index 1 since French is official) should be….**
  - **18% larger than those from Zambia to France (with a linguistic index 0.1)**
  - **6% larger that those from Sao Tome to France (with a linguistic index 0.7)**

Language proximity, other controls and ln. migration rates from 223 countries of origin to 30 OECD destination countries for 1980-2010.

VARIABLES	FE	Betas	VARIABLES Cont.	FE	Betas
	(8)	(9)		(8)	(9)
Linguistic Proximity	0.209***	0.020***	Ln Distance in km	-0.390***	-0.145***
	(0.066)			(0.030)	
Ln Stock of Migrants_t-1	0.669***	0.760***	Neighboring Dummy	-0.198**	
	(0.009)			(0.082)	
Ln Destination	1.723***	0.202***	Historical Past Dummy	0.261***	
GDPperCapPPPj_t-1	(0.132)			(0.092)	
Ln Origin	0.072	0.037	Dominant Genetic	0.00003	0.009
GDPperCapPPPi_t-1	(0.267)		Distance	(0.000)	
Ln Origin	-0.011	-0.097	Ln Origin Freedom	0.017	0.005
GDPperCapPPPit-1 squared	(0.016)		Political Rightsi_t-1	(0.023)	
Ln Destination	0.576***	0.056***	Ln Origin Freedom	-0.074***	-0.019***
Public Social Exp_t-1	(0.101)		Civil Rightsi_t-1	(0.028)	
Ln Destination	-0.051**	-0.010**	0/1 for Substiit. Unempl.	YES	YES
UnemplRate_t-1	(0.025)		Year, Dest & Origin FE	YES	YES
Ln Origin	0.054***	0.017***	Constant	-23.576***	
UnemplRate_t-1	(0.021)			(2.167)	
Ln Population Ratio_t-1	0.582***	0.550***	Observations	51,257	51,257
	(0.101)		Adjusted R-squared	0.899	0.899

Notes: Dependent Variable: Ln (Emigration Rate). Robust standard errors clustered at the country-pair level, *** p<0.01, ** p<0.05, * p<0.1.

To sum up

- Linguistic proximity important - Sharing the same language VS not sharing any level of the linguistic family tree has an effect on immigration flows equivalent to an increase of 12% in destination country GDP.

- The standardized beta-coefficients show:
- An increase in 1 st. dev. in stock of migrants is associated with a 0.76 st.dev. increase in migration rates. A similar increase in the income per capita (destination) increases migration to this country by 0.2 st.dev., whereas the implied impact of linguistic proximity is just a tenth of that, around 0.02 st.dev.
- The impact of having closer languages is larger than that of countries having higher (or lower) unemployment rates in origin (or destination) but less than half of the pull implied from larger social expenditures in destination.
-

Robustness: Additional linguistic variables

We recalculate all linguistic proximity indices

1. With language most extensively used in the country (sometimes not even official!)

   Ex: Angola Portuguese if 1st official among more than 6 officials but not the first or second most widely spoken; Philipinnes, Cebu most spoken and not official

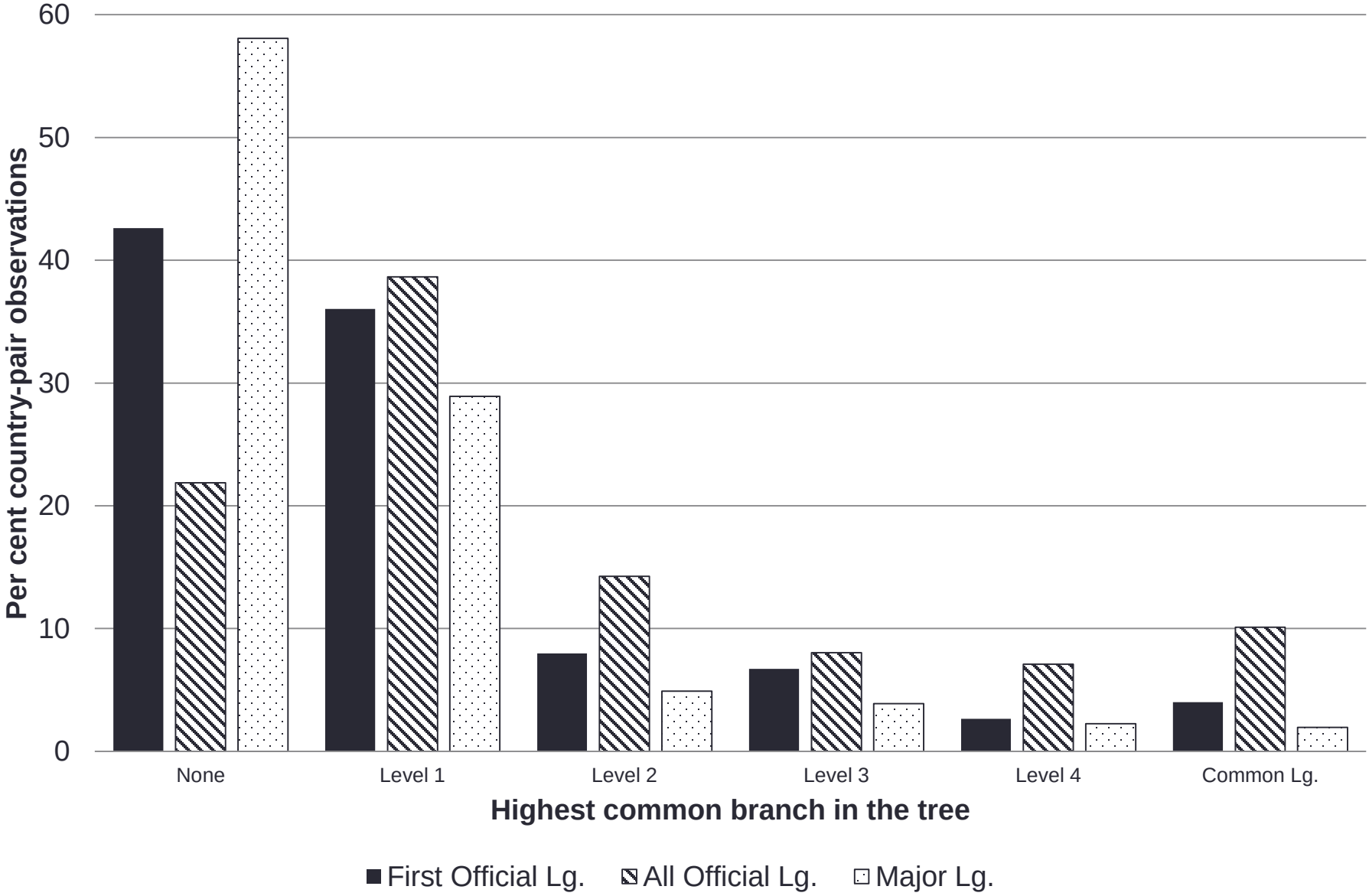
2. With the minimum distance between any of *multiple official languages and main languages spoken*

   Ex: Australia to Switzerland: Min distance from English to German, French, Italian or Romance

   Ex: India to Australia: min distance from English to either Hindi or English

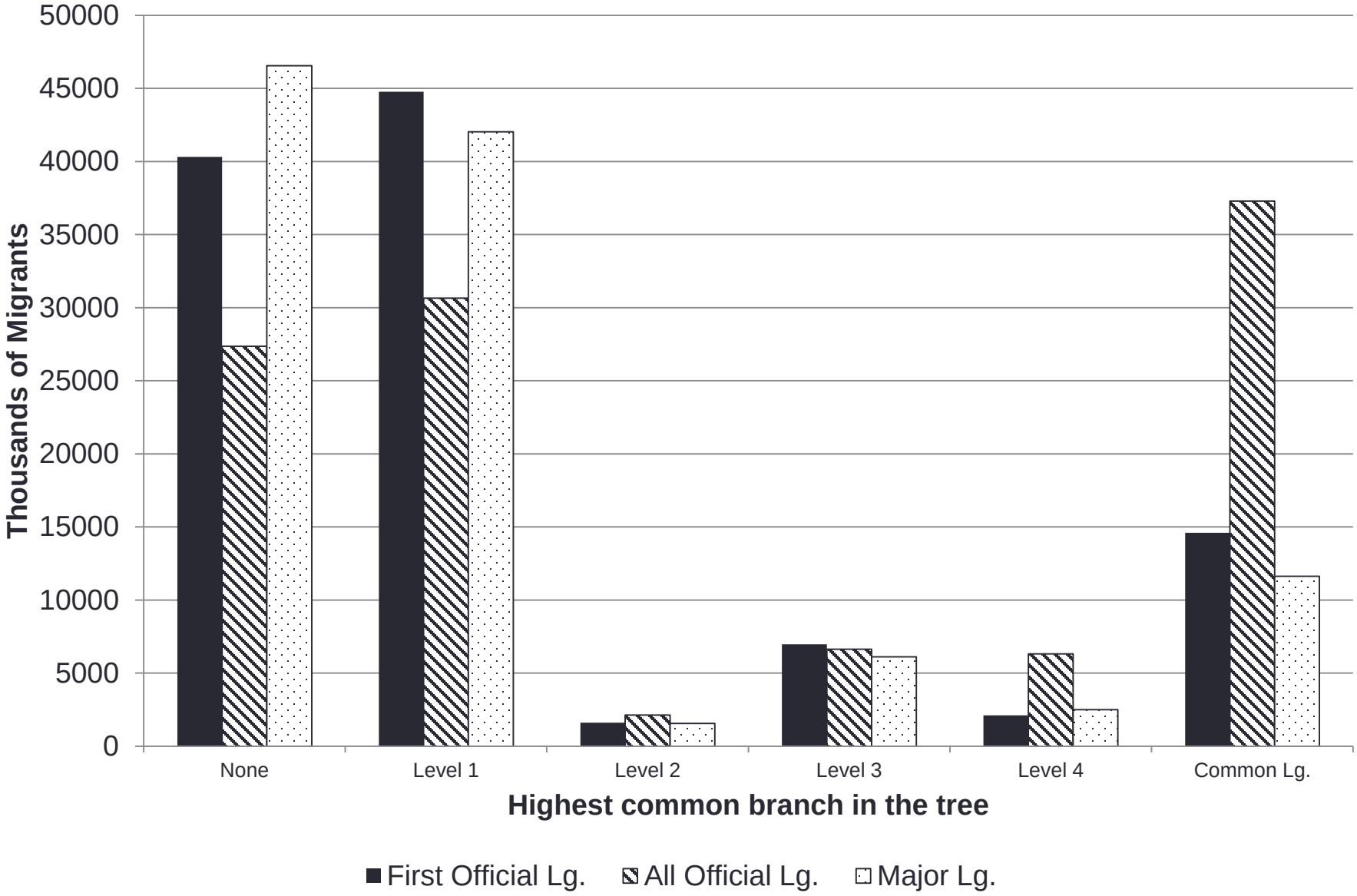
   Ex: Philipinnes to Australia: Tagale is 1st official and English 2nd official

Figure 1. Distribution of Country-pairs by Linguistic Proximity measured with Etnolinguistic Tree for 1980-2010



Unbalanced panel of 223 origin countries to 30 OECD destinations for period of 1980-2010

Figure 2. Migration Flows by Linguistic Proximity of countries measured with Ethnolinguistic Tree for 1980-2010

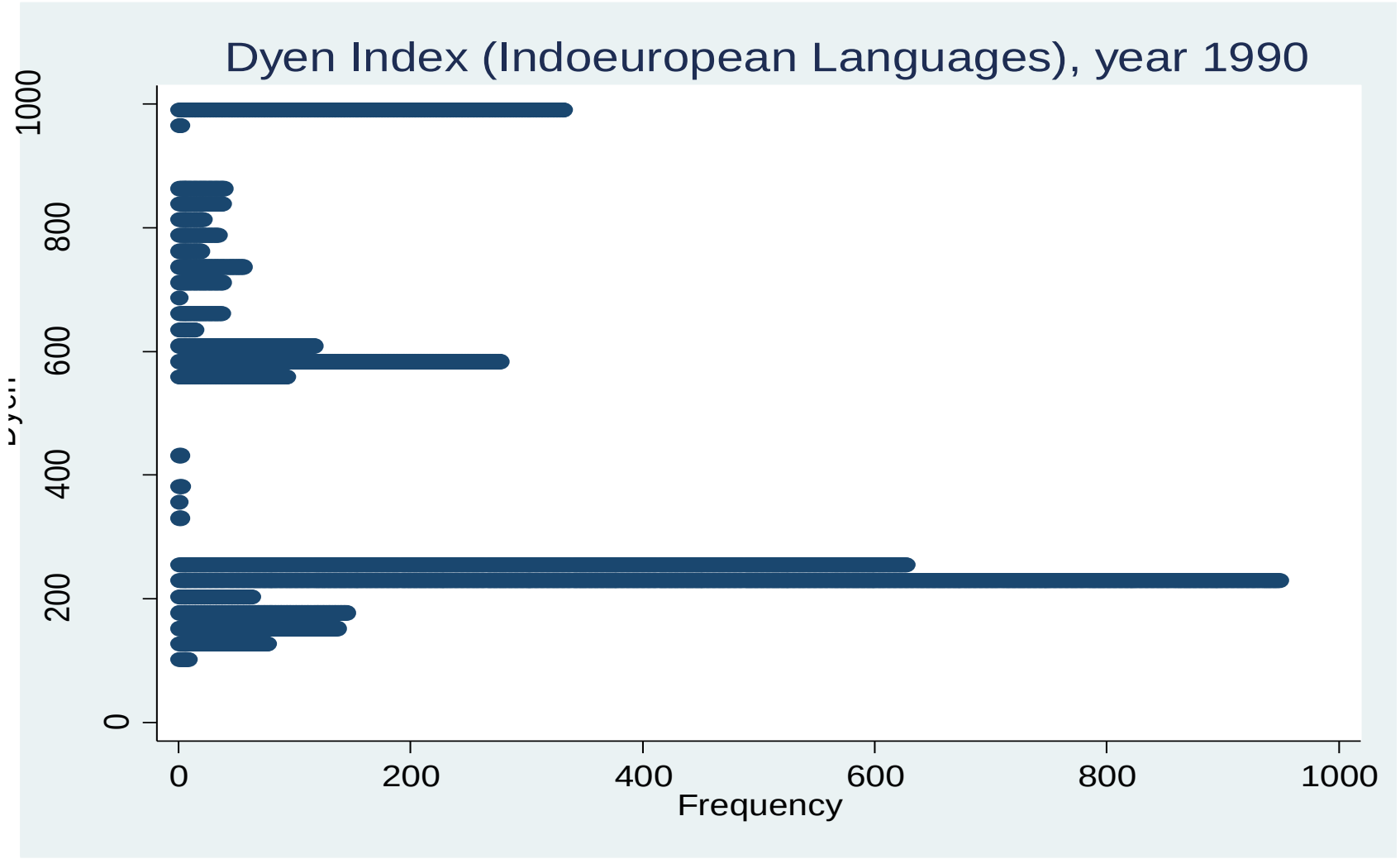


Robustness-Additional linguistic variables

Two continuous indices from linguists:

1. Proximity of Indo-European languages by Dyen et al. (1992), based on the proximity between samples of words (smaller sample size) (rescaled from 0 -1000 to 0-1 in estimates)

Dyen index (1000=equal language)



Robustness-Additional linguistic variables

Two continuous indices from linguists:

1. Proximity of Indo-European languages by Dyen et al. (1992), based on the proximity between samples of words (smaller sample size) (rescaled from 0 -1000 to 0-1 in estimates)
2. Distance which relies on phonetic dissimilarity of a core set of the 40 more common words across languages describing everyday life and items for all world languages, Levenshtein index developed in Max Planck institute.

Levenshtein index

Table 1: 40-Items Swadesh Word List

I	You	We	One
Two	Person	Fish	Dog
Louse	Tree	Leaf	Skin
Blood	Bone	Horn	Ear
Eye	Nose	Tooth	Tongue
Knee	Hand	Breast	Liver
Drink	See	Hear	Die
Come	Sun	Star	Water
Stone	Fire	Path	Mountain
Night	Full	New	Name

Source: Bakker et al. (2009).

Levenshtein index

- Words are expressed in a phonetic transcription and evaluated with the *ASJP code (Automatic Similarity Judgment Program)*
  - Ex: Mountain in English (maunt3n) to Berg in German (*bErk*).
- Finally compute the number of steps needed to move from one word expressed in one language to that same word expressed in the other language
  - This value is normalized to the maximum potential distance between two words. The sum of these distances is divided by number of words that exist in both compared lists and again nomalized by the similarity of phoneme inventories of the language pair. See Bakker et al (2009)
- In our sample from 0 (two languages are the same) to a maximum of 106.39 (for the distance between Laos and Korea).
- Defined as distance as opposed to the other indeces, thus we expect a negative sign.

Levenshtein index

	English	German	Steps
Fish	fis	fis	0
Breast	brest	brust	1
Hand	hEnd	hant	2
Tree	tri	baum	4
mountain	maunt3n	bErk	7

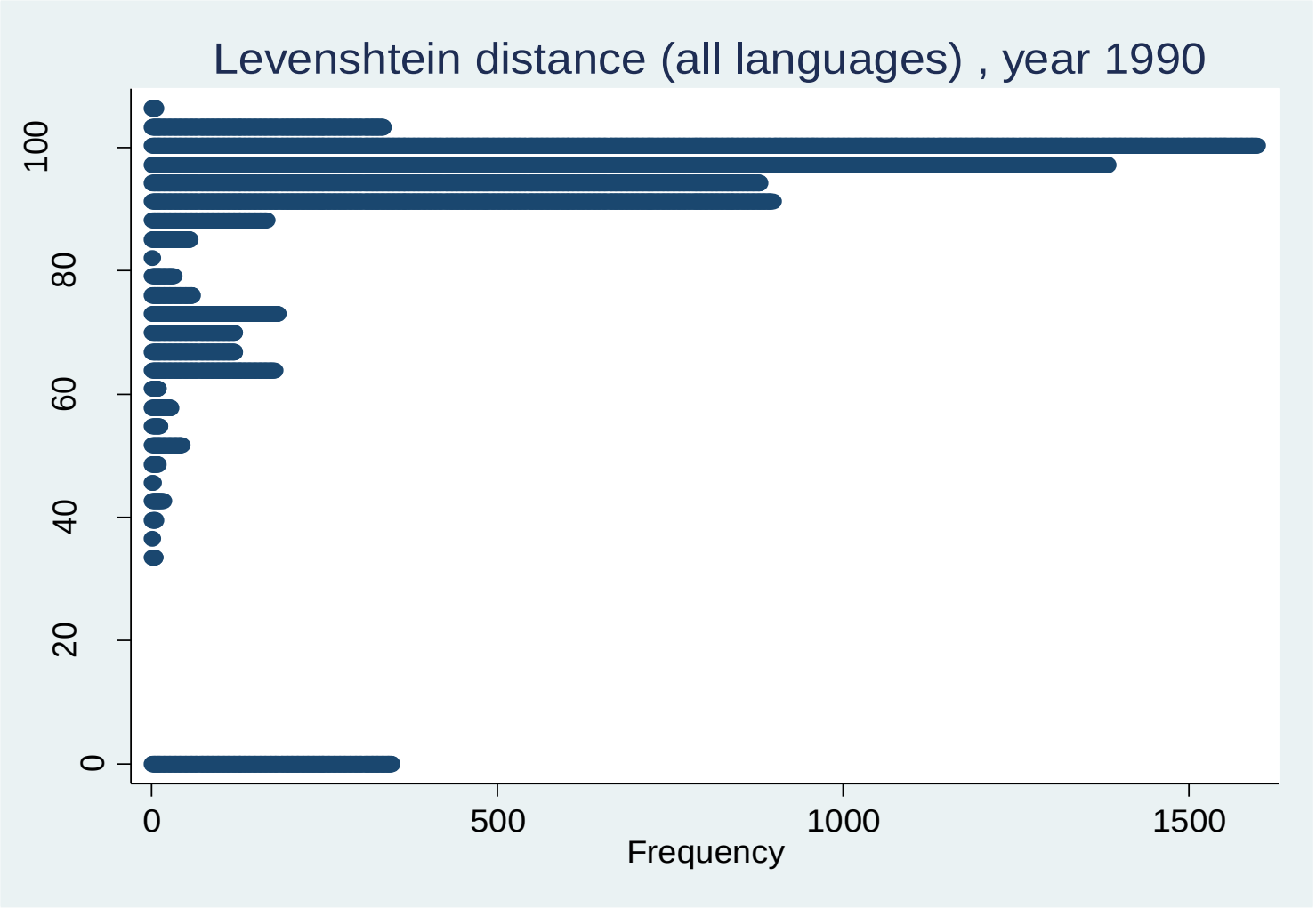
From Brown (2008); example used by Sinning (2013)

Levenshtein index - Closest and Furthest Languages

English				German			
Closest		Furthest		Closest		Furthest	
Language	*Distance*	*Language*	*Distance*	*Language*	*Distance*	*Language*	*Distance*
Dutch	63.22	Tamil	100.81	Swiss-German	48.34	Tamil	100.2
Norwegian	64.12	Turkish	101.04	Dutch	51.50	Hebrew	100.39
Swedish	64.40	Finnish	102.27	Norwegian	64.92	Indonesian	101.75
Danish	69.63	Somalian	103.03	Swedish	66.56	Malay	101.75
German	72.21	Vietnamese	104.06	Danish	66.96	Korean	104.3
French				**Czech**			
Closest		Furthest		Closest		Furthest	
Language	*Distance*	*Language*	*Distance*	*Language*	*Distance*	*Language*	*Distance*
Catalano	71.6	Irish	100.22	Slovak	32.59	Hebrew	99.55
Italian	73.89	Hungarian	100.65	Croatian	43.74	Vietnamese	99.72
Portuguese	74.36	Vietnamese	101.81	Serbian	43.74	Korean	99.85
Romanian	74.39	Japanese	101.94	Serbo-croatian	43.95	Chinese	101.12
Friulano	74.54	Korean	102.74	Polish	44.93	Japanese	101.76

Notes: – Source: Own calculations using programs for calculating ASJP distance matrices (Version 2.1),

Levenshtein index (0=equal language)



Comparing the three indices of linguistic distance - English

	Ethnologue	Dyen	Levenshtein
English-English	1	1000	0
English-Dutch	0.45	608	63.22
English-German	0.45	578	72.61
English-Spanish	0.1	240	98.03
English -Arabic	0	N/A	101.27

	Ethnol	Dyen	Levensh
Ethnologue	1.00		
Dyen	0.94	1.00	
Levenshtein	-0.93	-0.91	1.00

Robustness checks: alternative measures of linguistic proximity (Dyen, Levenshtein and controls for multiple official and main languages)

	First Official Language			All Official and Main Languages			Major Language		
Ling. Proximity/Distance measured by:	**Ling.Prox**	**Levensh.**	**Dyen**	**Ling.Prox**	**Levensh.**	**Dyen**	**Ling.Prox**	**Levensh.**	**Dyen**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Linguistic Proximity	0.209***	-0.144*	0.203***	0.192***	-0.199***	0.333***	0.355***	-0.218**	0.225**
	(0.066)	(0.076)	(0.077)	(0.054)	(0.058)	(0.066)	(0.085)	(0.099)	(0.096)
Z-score	**[0.020]*** **	**[-0.013]***	**[0.022]*** **	**[0.024]*** **	**[-0.023]*** **	**[0.039]*** **	**[0.027]*** **	**[-0.016]** **	**[0.023]** **
Observations	51,257	49,709	27,495	51,257	50,865	38,612	51,257	48,016	18,906

Similar relevance of linguistic proximity across all measures, around 20-15% higher migration rate from no linguistic similarity to complete in first official. Similar results using Dyen and Levenshtein.

Interpreting Levenshtein and Dyen coefficients

- Coeff -0,144 in col. (2) with Levenshtein (divided by 100):
  - emigration rates to countries with similar languages should be around 15% higher than to those with an index of around 100 (quite dissimilar).
- Coeff 0.203 in col. (3) with the Dyen index (divided by 1000):
- Emigration rates to an English speaking country like UK or US from Zambia (with a Dyen 1000 since English official) should be, ceteris paribus
  - Around 17% larger than from Nepal (with a Dyen of 157 with respect to English)
  - Around 15% larger than from Argentina (with an index of 240)
  - Around 8.5% larger than from Austria (with an index of 578)

Additional robustness: Separate dummies for coincidence at each level of linguistic tree

	(1)	(2)	(3)	(4)	(5)
Common Level 1	-0.032	-	-	-	-
	(0.069)	-	-	-	-
Common Level 2	-	0.125***	-	-	-
	-	(0.045)	-	-	-
Common Level 3	-	-	0.228***	-	-
	-	-	(0.047)	-	-
Common Level 4	-	-	-	0.345***	-
	-	-	-	(0.060)	-
Common Language	-	-	-	-	0.381***
	-	-	-	-	(0.091)
Ln Stock of Migrants_t-1	YES	YES	YES	YES	YES
Observations	26,235	26,235	26,235	26,235	26,235
Adjusted R-squared	0.876	0.876	0.876	0.877	0.876

Sharing the first level of the linguistic tree does not matter for migration flows

Sharing other levels of the linguistic tree matters incrementally

Additional robustness: Dummies for highest level of coincidence at tree for each pair

	(1)	(2)	(3)
Highest common linguistic Level:			
Level 1	0.183	0.235*	-0.055
	(0.140)	(0.129)	(0.072)
Level 2	0.602***	0.213	-0.112
	(0.169)	(0.156)	(0.086)
Level 3	0.426**	0.524***	0.021
	(0.179)	(0.161)	(0.092)
Level 4	1.246***	1.025***	0.234**
	(0.208)	(0.187)	(0.096)
Common (Level 5+)	1.751***	1.265***	0.360***
Year, origin & destination FE	YES	YES	YES
Economic & Political controls	NO	YES	YES
Lag Foreign Stock	NO	NO	YES
Observations	95,408	36,165	26,235
Adj. R2	0.620	0.751	0.877

Notes: Dependent Variable: Ln(Emigration Rate). Lagged dependent variable not included *** p<0.01, ** p<0.05, * p<0.1.

The role of widely spoken languages

- *Test whether the relevance of* linguistic proximity is similar for non-English speaking and for English-speaking destinations
- *Two different forces behind this:*

  1) Previous "proficiency" of English as second language because widely spoken (Internet, TV..), in business and taught at schools;

  2) English language proficiency is important skill, even at the labor market of source countries => learning/practicing/improving English attractive, especially for temporary migrants.

- *H:* If there is some advantage from knowing English, we expect that the linguistic proximity should matter more for non-English speaking destinations than for the others.

And education...

- research based on micro-data -2 polar types of migrants (see Belot and Hatton 2012; Docquier and Rappaport 2012 for an overview):
  - low skilled manual workers in jobs that are not filled by the natives in the destination country and,
  - high skilled professionals
- Language plays a key role in a skill transferability (Kossoudji, 1988; Bleakley and Chin, 2004; Chiswick and Miller, 2002, 2007, 2010; Dustmann, 1994; Dustmann and van Soest, 2001, 2002; and Dustmann and Fabbri, 2003) =>relevance of linguistic proximity and knowledge of widely spoken language will likely differ across various groups of migrants with different needs for skill transferability.
- *H:* linguistic proximity and knowledge of a widely spoken language are less relevant for migrants with lower average skills.

Table 5. The role of English as widely spoken language, education and migration rates to OECD countries.

	All countries			Countries with low levels of education		
	First Official	Major	All Official and Main	First Official	Major	All Official and Main
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic Proximity:						
In Non-English destination	0.363***	0.509***	0.225***	0.271*	-0.176	0.368***
	(0.073)	(0.082)	(0.059)	(0.144)	(0.287)	(0.099)
In English destination	0.061	0.108	0.150*	0.025	0.108	0.227**
	(0.095)	(0.147)	(0.083)	(0.123)	(0.237)	(0.100)
Obs	51,257	51,257	51,257	11,079	11,079	11,079

Less relevant for English Destinations

Table 5. The role of English as widely spoken language, education and migration rates to OECD countries, cont.

	First Official	First Official
	(7)	(8)
Linguistic Proximity:	0.244***	-0.014
	(0.067)	(0.126)
Origin Tertiary Education_t	0.109***	0.099***
	(0.022)	(0.022)
Linguistic Prox*Ter Edu_t		0.094**
		(0.043)
Other controls	YES	YES
Constant	-23.650***	-23.725***
	(2.210)	(2.208)
Observations	50,497	50,497
Adj. R2	0.899	0.899

Notes: Dependent Variable: Ln(Emigration Rate). A country with low education is below the 25th percentile in gross secondary school enrollment rates for a given year. Tertiary education is measured by gross enrollment rates. Controls included: stock of migrants, economic variables, distance variables, year dummies and destination and origin country fixed effects. Robust standard errors clustered at the country-pair level, *** p<0.01, ** p<0.05, * p<0.1.

The role of policy and linguistic networks

- Relevance of "Linguistic enclaves" (i.e. migrants for all Central America moving to highly Mexican areas in the US).
- Is the effect reinforced with linguistic proximity to the destination language?
- Language requirement as Immigration Policy
  - Difficult to measure in consistent way for entry
  - Easier to measure the requirement for naturalization (1 formal, 0.5 informal, 0 none)
    - *create a time-varying index that measures whether countries have any language requirement in the naturalization process - formal (i.e. written test) or informal and whether it has changed in each of the 30 OECD destinations for the 1980-2010 period*
    - *combine existing information from previous research (Goodman 2010a, Weil 2001, Waldrauch 2006, Joppke 2007), country official websites, data from the project EUDO Citizenship Observatory and legislation on citizenship by country available in the eudo-citizenship.eu.*

The role of policy and linguistic enclaves

	Policy (Naturalization)	Policy (Naturalization)	Linguistic networks at the 3rd level of the linguistic tree	Linguistic networks at the 4th level of the linguistic tree
Linguistic Proximity	0.205***	0.244**	0.311***	0.467***
Linguistic Requirement (Policy)_t 1 formal, 0.5 informal, 0 none)	-0.249***	-0.240***		
Ling.Req.Policy_t *Ling. Prox		-0.065		
Linguistic networks_t-1			0.040***	0.027**
Ling. Networks_t-1 *Ling. Prox			-0.035**	-0.065***
Ln Stock of Migrants_t-1	0.671***	0.671***	0.655***	0.661***
Constant	-23.374***	-23.374***	-23.847***	-23.770***
Observations	51,233	51,233	51,147	51,112

Findings on Control Variables

- Stock of migrants from same source: (+)
- Destination GDP p.c.: (+) weakens once unemployment included.
- Origin GDP p.c. (nonlinear).
- Unemployment (scarce data; + at origin; – at destination)
- Public social expenditure at destination (+) (“welfare magnet”?)
- Distance (-), Colonial Past (+),
- Restrictive political rights at origin (-), restrictive civil rights (seem +, not robust)

Summary I

- Migration flows between countries with the same 1st official language compared to those with no similarity at any level of the linguistic family tree are around 20% larger, ceteris paribus.
- Robust to:
  1. Use multiple official and main languages or most widely used language in the country
  2. Continuous distance measures of IndoEuropean languages (Dyen) or of all world countries (Levenshtein)
  3. Inclusion of Genetic distance
- In the context of traditional economic push & pulls, the impact of linguistic proximity is lower than that of ethnic networks or destination GDP per capita level, but stronger than that of unemployment rates.

Summary II

- Linguistic proximity stronger predictor of migration flows for non-English speaking destinations. Less relevant for migrants coming from countries with low levels of education.
- Migration flows are smaller in countries with higher linguistivc policy requirements, but the relevance of linguistic proximity remains unaltered
- migration rates are larger in destinations with larger size of the linguistic community, where the pressure to learn the local language immediately after arrival is likely to be lower. Our estimates reveal that the linguistic proximity matters less when the size of the linguistic community is large in destinations.

Other Research

- **Apply linguistic distance indices to micro-data to study socio-economic outcomes and adaptation of migrants to new environment.**

- **Apply migration dataset for a number of projects (till now: climate, natives attitudes, immigrant rights, welfare magnet, relative deprivation, studies of consequences of migration….)**

Swadesh 100-item list (Swadesh 1971: 283)

1. I	21. dog	41. nose	61. die	81. smoke
2. you	22. louse	42. mouth	62. kill	82. fire
3. we	23. tree	43. tooth	63. swim	83. ash
4. this	24. seed	44. tongue	64. fly	84. burn
5. that	25. leaf	45. claw	65. walk	85. path
6. who	26. root	46. foot	66. come	86. mountain
7. what	27. bark	47. knee	67. lie	87. red
8. not	28. skin	48. hand	68. sit	88. green
9. all	29. flesh	49. belly	69. stand	89. yellow
10. many	30. blood	50. neck	70. give	90. white
11. one	31. bone	51. breasts	71. say	91. black
12. two	32. grease	52. heart	72. sun	92. night
13. big	33. egg	53. liver	73. moon	93. hot
14. long	34. horn	54. drink	74. star	94. cold
15. small	35. tail	55. eat	75. water	95. full
16. woman	36. feather	56. bite	76. rain	96. new
17. man	37. hair	57. see	77. stone	97. good
18. person	38. head	58. hear	78. sand	98. round
19. fish	39. ear	59. know	79. earth	99. dry
20. Bird	40. Eye	60. sleep	80. cloud	100. name

Robustness: including dummy for common language & linguistic distance together in the FE model

	(1)	(2)	(3)
Linguistic Proximity	0.436***		0.353***
	(0.081)		(0.098)
Common Language		0.381***	0.122
		(0.090)	(0.112)
Unemployment rates	YES	YES	YES
Year, origin and destination FE	YES	YES	YES
Observations	26,235	26,235	26,235
Adj. R2	0.877	0.877	0.877

Notes: Dependent Variable: Ln(Emigration Rate). Controls included: stock of migrants, economic variables, distance variables,. Lagged dependent variable not included *** p<0.01, ** p<0.05, * p<0.1.

Language proximity and ln. migration rates from 223 countries of origin to 30 OECD destination countries for 1980-2010.

Δ in St. Dev migration rates from Δ one St Dev (betas)

	OLS	OLS		FE	Beta	Poisson
VARIABLES	(1)	(2)	(3)	(4)		(5)
Linguistic Proximity	3.271***	-	3.343***	0.209***	0.020***	0.508***
	(0.147)		(0.215)	(0.066)		(0.121)
Common Language	-	2.929***	-0.095	-		
		(0.169)	(0.254)			
Ln Stock of Migrants_t-1	NO	NO		YES	YES	YES
Unemployment rates	NO	NO		Subs	Subs	Subs
Destination & Origin FE	NO	NO		YES	YES	YES
Observations	100.519	100,519		51,257	51,257	51,257
Adjusted R-squared	0.111	0.076		0.863	0.863	0.902

↑R^2 with proximity index

Notes: Dependent Variable: Ln (Emigration Rate). Controls included: **stock of migrants, economic & political variables, distance variables, colonial, year dummies and destination and origin country fixed effects**. Robust standard errors clustered at the country-pair level, *** p<0.01, ** p<0.05, * p<0.1.

Interpretation 1980-2010

- Cols (3): Emigration flows to a country with same language as opposed to one with no common family should be around 20% higher.
- When comparing emigration rates to France **in (3):**
- **Ceteris paribus, rates from Benin (with index 1 since French is official) should be….**
  - **18% larger than those from Zambia to France (with a linguistic index 0.1)**
  - **6% larger that those from Sao Tome to France (with a linguistic index 0.7)**

Ln migration rates: Alternative Linguistic Measures

	First Official Language	
Ling. Proximity/Distance measured by	Levenshtein (All countries Phonetic similarity)	Dyen (Indo-European Word similarity)
	(1)	(2)
Linguistic Proximity/Distance	0.4***	0.4***
	(0.001)	(0.000)
Observations	25,770	15,301
Adj. R2	0.875	0.872

Similar relevance of linguistic proximity across all measures, around 40% (in sample without substituted unemployment) higher migration rate from no linguistic similarity to complete.

Interpreting Dyen coefficient

- Emigration rates to an English speaking country like UK or US from Zambia (with a Dyen 1000 since English official) should be, ceteris paribus (in models without lagged dependent)
  - Around 34% larger than from Nepal (with a Dyen of 157 with respect to English)
  - Around 30% larger than from Argentina (with an index of 240)
  - Around 17% larger than from Austria (with an index of 578)

Comparing the three indices of linguistic distance normalized to z-scores

	Ethnologue	Dyen	Levenshtein
Linguistic Proximity	0.068***	0.078***	0.057***
	(0.017)	(0.023)	(0.018)
Unemployment rates	NO	NO	NO
Observations	47,910	25,083	46,558
Adj. R2	0.877	0.877	0.862
Linguistic Proximity	0.109***	0.121***	0.095***
	(0.020)	(0.026)	(0.0215)
Unemployment rates	YES	YES	YES
Observations	26,235	15,301	25,770
Adj. R2	0.877	0.877	0.876

Notes: Dependent Variable: Ln(Emigration Rate). Controls included: stock of migrants, economic variables, distance variables, year dummies and destination and origin country fixed effects. Lagged dependent variable not included *** p<0.01, ** p<0.05, * p<0.1.

Ln migration rates: Alternative Linguistic Measures

	All Official Languages			Major Language		
Ling. Proximity/Distance measured by	Ling.Proximity	Levenshtein	Dyen	Ling.Proximity	Levenshtein	Dyen
Linguistic Proximity/Distance	0.368*** (0.071)	0.40*** (0.001)	0.10*** (0.000)	0.481*** (0.089)	0.40*** (0.001)	0.50*** (0.000)
Observations	26,235	26,180	19,970	26,235	25,841	13,170
Adj. R2	0.876	0.877	0.877	0.837	0.875	0.872

Notes: Dependent Variable: Ln (Emigration Rate). Controls included: stock of migrants, economic & political variables, distance variables, colonial, year dummies and destination and origin country fixed effects. Robust standard errors clustered at the country-pair level, *** p<0.01, ** p<0.05, * p<0.1.

Similar relevance of linguistic proximity across all measures, around 37-40% higher migration rate from no linguistic similarity to complete . Similar results using Dyen and Levenshtein. Sample with no substituted unemployment.

Robustness: missing unemployment

- We have re-run all models substituting missing unemployment rates observations for a country by the average unemployment in that country
- Results do not substantially change even if sample size increases from around 26,000 to 51,000. The coefficient for linguistic proximity is similar as when unemployment rates are not included in model and the sample is larger.

Robustness : Adding controls for Genetic Distance

- Add to the model two indices of genetic distance
  - Measure distance of distributions of alleles in both populations by Cavalli-Sforza, Menozzi, and Piazza 1994) and takes value 0 for identical.
  - **dominant**: distance between the plurality ethnic groups of each country in a pair (=the groups with the largest shares of each country's population)
  - **weighted:** using all existing groups, expected genetic distance between two randomly selected individuals, on from each country.
- Purpose: To rule out that language is masking other factors such as cultural or genetic similarity among populations.
- Findings: No change in size and significance of coefficients of linguistic distance

Adding controls for Genetic Distance

Linguistic Proximity measured by:	No Linguistic Variable	No Linguistic Variable	Linguistic Proximity	Linguistic Proximity
Linguistic Proximity	NO	NO	0.462***	0.458***
			(0.082)	(0.082)
Dominant Genetic Distance	0.000	-	0.000**	-
	(0.000)	-	(0.000)	-
Weighted Genetic Distance	-	0.000	-	0.000
	-	(0.000)	-	(0.000)
Constant	-23.744***	-23.680***	-24.153***	-24.082***
	(3.431)	(3.427)	(3.450)	(3.448)
Observations	26,136	26,014	26136	26014
Adj. R2	0.876	0.876	0.877	0.877

Note: When the sample is restricted to Indo-European countries (more homogenous) the sign of the genetic distance is negative as expected though only significant for weighted. Thus for relatively closer countries genetics matter more to explain migration flows than when we look at the complete sample of the world.

Genetic distance not significant

The Role of Widely spoken languages (English vs Non English destinations)

	First Official Language		Major Language	All Official Languages
	(1)	(2)	(3)	(4)
Linguistic Proximity:				
In Non-English destination	0.538***	0.409***	0.620***	0.294***
	(0.082)	(0.077)	(0.086)	(0.077)
In English destination	0.283**	0.126	0.219	0.479***
	(0.141)	(0.106)	(0.175)	(0.112)
Ln Emigration Rate_t-1	NO	NO	NO	NO
Other controls	YES	No Unempl. rates	YES	YES
Observations	26,235	47,910	26,235	26,235
Adj. R2	0.877	0.863	0.877	0.876

Less relevant for English Destinations

More relevant when multiple official languages taken into account

Even Less relevant when more heterogenous origin countries added
Note larger sample size

OUR NEXT LECTURE – Monday 18.1.2016

- ***Determinants of migration II***
- ***Selectivity in migration, models of migration and empirical evidence***

THE NEXT LECTURES

- *Immigrant performance and integration; the second generation*
- *Immigrants and innovation; International migration and globalization*
- *Impacts of immigration*
- *Immigration policy*
- *Diversity - Impacts of workforce diversity on firms and economies*
- *Emigration and source countries; Brain drain and brain gain; Remittances*