Quaderni del Premio «Giorgio Rota»

N. 4, 2016

The Economics of Migration

Centro di Ricerca e Documentazione Luigi Einaudi

Con il sostegno di Fondazione CRT

Centro di Ricerca e Documentazione Luigi Einaudi

Quaderni del Premio «Giorgio Rota»

n. 4, 2016

The Economics of Migration

Iniziativa realizzata con il sostegno di

Fondazione CRT

Gli autori di questo Quaderno:

Ainhoa Aparicio Fenoll
Assistant Professor, Collegio Carlo Alberto, Turin (Italy)

Simone Bertoli
Associate Professsor, CERDI, Université d'Auvergne (France)

Zoë Kuehn
Assistant Professor Universidad Autónoma de Madrid (Spain)

Ilse Ruyssen
BOF postdoctoral research fellow, SHERPPA, Ghent University (Belgium)

Xingna Nina Zhang
PhD student and a part-time RA in School of Geographical Sciences,
University of Bristol (United Kingdom)

Alessandra Venturini

Professore Ordinario Università di Torino, Deputy Director Migration Policy Center, European University Institute, Membro del Direttivo di FIERI

Centro di Ricerca e Documentazione Luigi Einaudi
Via Ponza 4 • 10121 Torino • segreteria@centroeinaudi.it
www.centroeinaudi.it

Quaderni del Premio «Giorgio Rota»
n. 4, 2016

Indice

Il Premio «Giorgio Rota»

L'intento del Premio *«Giorgio Rota» Best Paper Award* è di riprendere l'attività di ricerca annualmente condotta dal Comitato / Fondazione Giorgio Rota prima della sua inclusione nel Centro Einaudi, sulla relazione tra il pensiero e l'agire economico e un aspetto (ogni anno diverso) del vivere in società, mantenendo vivo il ricordo e l'insegnamento dell'economista Giorgio Rota, uno dei primi animatori del Centro, prematuramente scomparso.

Dal 2012 il Cento Einaudi ha dunque raccolto questa eredità rinnovando la formula della ricerca: è stato perciò istituito questo premio annuale dedicato a giovani ricercatori, con una qualificazione accademica nei campi dell'economia, sociologia, geografia, scienza politica o altre scienze sociali. I paper possono essere presentati sia in italiano che in inglese, e non devono essere stati pubblicati prima della data della Conferenza Rota, l'evento pubblico nel quale i vincitori hanno modo di presentare il loro lavoro.

La prima edizione 2012 aveva per tema *Contemporary Economics and the Ethical Imperative* e la Conferenza Giorgio Rota 2013/Giorgio Rota Conference 2013, si è tenuta presso il Centro Einaudi il 25 marzo 2013 con keynote speech di Alberto Petrucci, LUISS Guido Carli, Roma.

La seconda edizione, nel 2013, è stata su *Creative Entrepreneurship and New Media* con Conferenza Giorgio Rota 2014 presso il Centro Einaudi, 14 aprile 2014 e keynote speech di Mario Deaglio, Università di Torino.

La terza edizione, del 2014, ha analizzato il tema *The Economics of Illegal Activities and Corruption*, con Conferenza Giorgio Rota 2015 presso il Centro Einaudi, 15 giugno 2015. Keynote speech di Friedrich Schneider, Johannes Kepler University (Linz, Austria).

La quarta edizione, 2016, verteva su *The Economics of Migration*. Il 20 giugno 2016 si è tenuta la Conferenza Giorgio Rota 2016, presso il Campus Luigi Einaudi. Keynote speech di Alessandra Venturini, Università di Torino.

I paper vincitori della quarta edizione del Premio, Ainhoa Aparicio Fenoll e Zoë Kuehn, Simone Bertoli e Ilse Ruyssen e Xingna Nina Zhang sono riportati in questo volume.

Chi era Giorgio Rota

Giorgio Rota (1943-1984) è stato professore di Economia politica presso l'Università di Torino, e consulente economico. Per il Centro Einaudi, è stato coordinatore agli studi e membro del comitato di direzione di «Biblioteca della libertà».

Le sue pubblicazioni scientifiche abbracciano diversi temi: l'economia dei beni di consumo durevoli, l'economia del risparmio, il mercato monetario e finanziario, l'inflazione e la variazione dei prezzi relativi, il debito pubblico. Ricordiamo tra esse: *Struttura ed evoluzione dei flussi finanziari in Italia: 1964-73* (Torino, Editoriale Valentino, 1975); *L'inflazione in Italia 1952/1974* (Torino, Editoriale Valentino, 1975); nei «Quaderni di Biblioteca della libertà», *Passato e futuro dell'inflazione in Italia* (1976) e *Inflazione per chi?* (1978); *Che cosa si produce come e per chi. Manuale italiano di microeconomia*, con Onorato Castellino, Elsa Fornero, Mario Monti, Sergio Ricossa (Torino, Giappichelli, 1978; seconda edizione 1983); *Investimenti produttivi e risparmio delle famiglie* (Milano, Il Sole 24 Ore, 1983); *Obiettivi keynesiani e spesa pubblica non keynesiana* (Torino, 1983).

Tra le sue ricerche va particolarmente citato il primo *Rapporto sul risparmio* e sui risparmiatori in Italia. Rilevazione relativa all'anno 1982, risultato di un'indagine sul campo condotta da BNL-Doxa-Centro Einaudi, le cui conclusioni riscossero notevole attenzione da parte degli organi di stampa. Da allora il *Rapporto sul risparmio* continua a essere pubblicato ogni anno.

Quaderni del Premio «Giorgio Rota»
n. 4, 2016

Alessandra Venturini

Presentazione

Ogni giorno i giornali e la televisione mostrano immagini di migranti, di persone che parlano di migranti, di amministratori che gestiscono l'emergenza del fenomeno migratorio o di politici che discutono, commentano, criticano e tentano di prendere decisioni su tale fenomeno.

Il pubblico, di tutto questo spettacolo, è sempre più confuso e lacerato da informazioni contrastanti e largamente emotive. L'informazione quotidiana diventa disinformazione quotidiana perché molto raramente viene ricostruito il contesto e solo la notizia ha risalto.

Per questo motivo quest'anno il Centro Einaudi ha voluto dedicare il Premio «Giorgio Rota» / Giorgio Rota Best Paper Award al tema della migrazione, con la pubblicazione dei tre migliori papers *di giovani ricercatori su «The economics of migration», in modo da contribuire concretamente alla ricerca scientifica, la sola che può fornire risposte e aiutare a capire le dinamiche del fenomeno migratorio.*

I tre lavori sono molto differenti ma ugualmente caratterizzati da rigore e chiarezza e in modi diversi focalizzati a capire e spiegare la scelta migratoria.

Il primo, di Ainhoa Aparicio Fenoll e Zoë Kuehn, analizza il ruolo dell'istruzione – e, in particolare, dell'arricchimento della formazione con l'apprendimento di una lingua straniera – nel frenare o favorire la migrazione. L'investimento in istruzione è sempre percepito come positivo per un paese perché aumenta il capitale umano dei suoi cittadini e può favorire la crescita sia economica che sociale. La maggiore istruzione è diventata la carta vincente per emigrare, da ciò nasce l'annoso dibattito sul brain drain *che rappresenta uno dei dilemmi attuali della politica di istruzione e della politica migratoria. Per la prima volta uno studio empirico cerca di quantificare l'effetto della maggiore istruzione, specificando il contenuto dell'insegnamento. La ricerca dei due autori mostra che se la crescita dell'istruzione, in generale, frena la migrazione – perché gli individui, essendo più istruiti, hanno a disposizione più opportunità di lavoro – l'insegnamento di una lingua straniera, aumentando il campo delle opportunità lavorative, favorisce invece la migrazione. Ipotizzare di calmierare l'insegnamento delle lingue straniere per timore di una fuga verso l'estero degli studenti non è certo una soluzione efficiente, perché la capacità di un paese di attirare investimenti dall'estero dipende anche dalla possibilità di comunicazione dei suoi lavoratori. Se dunque nel breve periodo la fuga dei cervelli può rappresentare un costo, l'importante è nel medio-lungo periodo saper attirare investimenti.*

Il secondo paper*, di Simone Bertoli e Ilse Ruyssen, utilizza i dati Gallup World Pulls per analizzare il ruolo giocato dai* network *familiari dello stesso gruppo etnico nel favorire la migrazione. Occorre tener presente che i dati Gallup sulla "propensione ad emigrare" devono essere utilizzati con grande cautela nel prevedere i futuri flussi migratori, perché da un lato le intenzioni spesso non danno origine a veri trasferimenti e dall'altro il disegno del campionamento è realizzato per fornire indicazioni e non numeri programmati. Tuttavia, essi contengono informazioni preziose sui meccanismi di scelta della migrazione: la catena migratoria ha da sempre giocato un ruolo fondamentale nella scelta della destinazione, condizionandone i costi e la disponibilità informativa. Gli autori cercano di stimare come la presenza di amici o familiari in una potenziale destinazione condizioni la scelta verso tale area.*

Il terzo paper *si riallaccia a una esperienza ben nota al contesto italiano e mostra come le dinamiche della migrazione siano le stesse di un contesto interno e poi si adattino a un contesto internazionale. Xingna Nina Zhang analizza le dinamiche rurali-urbane della migrazione cinese e gli spostamenti di popolazione tra province nel decennio 2000-2010. La migrazione nasce come uno spostamento della popolazione che si muove dove ci sono più opportunità di lavoro e di reddito; i costi della migrazione interna sembrano minori di quelli necessari per la migrazione internazionale ma la distanza linguistica, climatica e ambientale sovente non sono da meno. Stimare la mobilità interna cinese vuole dire poterla poi anche prevedere e, data l'importanza della Cina quale* player *internazionale, diventa estremamente rilevante per capire se i flussi migratori internazionali cresceranno spontaneamente o se invece l'invecchiamento della popolazione porterà a un'inversione di direzione, con la crescita dell'immigrazione dall'estero verso la Cina. Come sottolineato, l'importanza della Cina sia come investitore nei paesi in via di sviluppo (per esempio in Africa), sia in genere nel mondo occidentale, la rende cointeressata ai problemi che l'Europa affronta in questo periodo con la crescita delle richieste di protezione internazionale. Come l'ultimo G20 a Pechino (settembre 2016) sottolineava, il tema della migrazione deve essere affrontato in modo multilaterale perché le politiche adottate da un paese in tema di riconoscimento dell'asilo hanno un effetto immediato sui paesi vicini, creando tensioni o risolvendo emergenze.*

L'intento perseguito dal Centro Einaudi, attraverso la quarta edizione del Premio «Giorgio Rota», è stato di premiare i tre articoli sopra illustrati e farli presentare dai giovani autori durante la Conferenza Giorgio Rota 2016, per contribuire alla conoscenza del tema migratorio con la diffusione di un'informazione scientifica.

Simone Bertoli
Ilse Ruyssen

Networks and migrants' intended destination[1]

Abstract. Social networks are known to influence migration decisions, but connections between individuals can hardly be observed. We rely on individual-level surveys conducted by Gallup in 147 countries that provide information on migration intentions and on the existence of distance-one connections for all respondents in each of the potential countries of intended destination. The origin-specific distribution of distance-one connections from Gallup closely mirrors the actual distribution of migrant stocks across countries, and bilateral migration intentions appear to be significantly correlated with actual flows. This unique data source allows estimating origin-specific conditional logit models that shed light on the value of having a friend in a given country on the attractiveness of that destination. The validity of the distributional assumptions that underpin the estimation is tested, and concerns about the threats to identification posed by unobservables are substantially mitigated.

Keywords. International migration; networks; intentions

1. Introduction

Social networks are expected to exert a key influence on migration decisions: connections with individuals that have already moved contribute to improve job prospects at destination (Munshi 2003; Patel and Vella 2013) and they can reduce the multifaceted costs of crossing a border (Carrington *et al.* 1996), while networks at origin can reduce the incentives to move (Munshi and Rosenzweig 2016). The existing empirical evidence on the effects of networks at destination on migration is based on rather coarse measures of networks, such as the share of households with a migrant at the village (McKenzie and Rapoport 2010) or at the county level (Bertoli 2010), or the size of the diaspora in each destination country (Pedersen *et al.* 2008; Beine *et al.* 2011, 2015; Beine and Salomone 2013; Bertoli and Fernández-Huertas Moraga 2015). The implicit

[1] The authors are grateful to Frédéric Docquier, Francesco Fasani, Jesús Fernández-Huertas Moraga, Elisabetta Lodigiani, Joël Machado, Elie Murard, Hillel Rapoport, Steven Stillman and to the participants to the International Migration Workshop at CERDI, to the 5th Meeting of Belgian Economists at Louvain-La-Neuve, to the 2nd Workshop on the Economics of Migration at Frankfurt, the Barcelona GSE Migration Workshop and the 9th Conference on Migration and Development for their comments, to Robert Manchin and the Gallup Institute for Advanced Behavioural Studies for providing access to the data for the purpose of this project, and to Olivier Santoni for providing research assistance; the authors also gratefully acknowledge the nancial support of the Centro Einaudi through the Giorgio Rota Prize; Simone Bertoli acknowledges the support received from the Agence Nationale de la Recherche of the French government through the program "Investissements d'avenir" (ANR-10-LABX-14-01); the usual disclaimers apply.

assumption behind this approach, which reflects binding data constraints, is that all potential migrants equally benefit from the networks at destination.[2] This assumption is at odds with theoretical representations of social networks (Jackson 2010) and with the empirical evidence on how members of a migrant network interact with each other (Comola and Mendola 2015).

Our objective is to contribute to gaining a deeper understanding of how social networks influence international migration by using a dataset that provides unique information on the individual-level connections to networks in each potential destination. Specifically, we draw on the data from 419 surveys conducted by Gallup in 147 countries of the world between 2007 and 2011 (Gallup 2013). For each respondent, we have information on whether she has relatives or friends who reside abroad, as well as on the countries in which they reside.[3] Reassuringly, the geographical distribution of distance-one connections for each country closely matches the actual bilateral distribution of migrants across destinations for 2010.

We combine the information on the countries in which a respondent has a distance-one connection with information on whether she intends to migrate and, if this is the case, to which destination. The Gallup World Polls do not provide information about actual moves, but we provide econometric evidence that the bilateral number of intending migrants by year is significantly associated with the yearly scale of actual bilateral migration flows to OECD destinations.[4]

A few studies have so far relied on the Gallup World Polls to investigate the patterns and determinants of migration intentions, without using the information about the preferred destination. Specifically, Esipova *et al.* 2011 present a detailed descriptive analysis of migration intentions; Manchin *et al.* 2014 analyze the effect of individual satisfaction on the desire to migrate, while Dustmann and Okatenko 2014 evidence that the relationship between the intention to move (either internally or across borders) and wealth is non-monotonic. Docquier *et al.* 2015 and Delogu *et al.* 2015 have used the origin-specific proportion of the individuals who intend to move to each foreign destination in their analyses of the short- and long-run efficiency gains of a removal of the legal restrictions to migration, assuming that the answers to the hypothetical questions in the Gallup World Polls are informative about the scale of liberalized migration flows. Docquier *et al.* 2014 empirically analyze the country-specific and dyadic factors governing the size and the composition of the bilateral pool of intending migrants, as well as the probability that these intentions are realized.

[2] The estimation of gravity equations derived from underlying random utility maximization models on aggregate data has to rest on this assumption, as the equivalence of the estimates obtained on aggregate and on individual-level data depends on the absence of individual-specific regressors (Guimaraes *et al.*, 2003); Munshi (2016) reviews additional concerns related to the identication of network effects from gravity equations on aggregate data on bilateral migration flows.

[3] This destination-specific dimension of the information is what distinguishes the data that we use from the dataset on internal Chinese migration used by Giulietti *et al.* (2014), who have information about whether each individual has a friend residing in an (unspecified) Chinese urban area.

[4] Creighton (2013), Dustmann and Okatenko (2014), Chort (2014), Manchin *et al.* (2014) and Docquier *et al.* (2014) also provide empirical evidence on the relationship between stated intentions and actual migration.

We estimate, separately for each of the 147 countries in our sample, a conditional logit model that describes the choice of intending migrants among the alternative destinations and that controls for the dependency of location-specific utility on the size of the diaspora. The estimation reveals that having a distance-one connection in a country is, on average, associated with an increase in the relative odds of opting for that destination by six to eight times, conditional upon intending to migrate. Distance-one connections have a relatively small effect compared to the dispersion in the deterministic component of location-specific utility of all countries in the choice set that are implied by our estimates, but main destinations are characterized by a similar level of attractiveness, so that distance-one connections can tilt the balance among them.

Our estimation approach is exposed to the threats to identification posed by correlated peer effects, i.e., unobserved factors that influence both the geographical distribution of one's own peers and the attractiveness of the various potential destinations, which would also jeopardize the distributional assumptions that justify the estimation of a conditional logit model. We follow two distinct and complementary approaches to address the concerns that our evidence about the key role played by distance-one connections in determining the preferred intended destinations is just reflecting correlated peer effects.[5] Specifically, we (*i*) add further individual-level variables drawn from the Gallup World Polls, and (*ii*) re-estimate the model on suitably restricted choice sets. Although we cannot fully dismiss the concerns related to the effects of unobservables on our estimates, the results from the various alternative specifications that we bring to the data greatly help to substantially mitigate them. The remainder of the paper is structured as follows. Section 2 introduces the data from the Gallup World Polls. Section 3 briefly describes the random utility model that describes the location-decision problem that intending migrants face. Section 4 contains some basic descriptive statistics, and Section 5 presents the benchmark estimates, and it discusses a number of threats to identification. Finally, Section 6 draws the main conclusions.

2. The Gallup World Polls

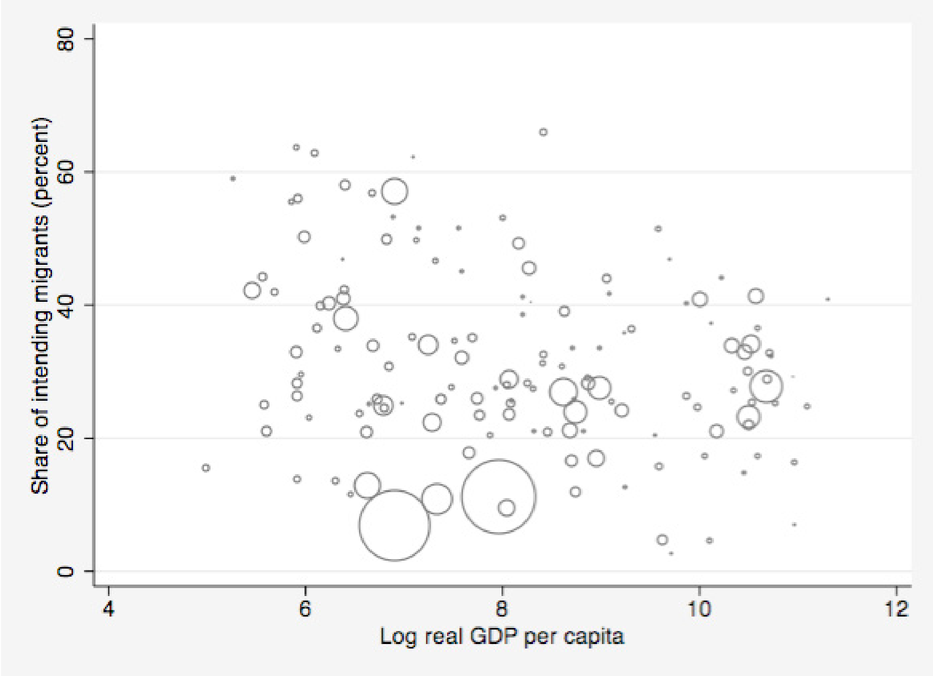
Our analysis rests on individual-level data from 147 countries where at least one Gallup World Poll has been conducted between 2007 and 2011.[6] The surveys conducted by Gallup typically have a sample of around 1,000 randomly selected respondents per country, and the data are collected either through face-to-face interviews or through phone calls in countries where at least 80 percent of the population has a telephone land-line.

[5] The Gallup World Polls do not provide information on the entire network, so that we do not have information on the geographical distribution of distance-two connections, which might have otherwise been used in the estimation to correct for the possible endogeneity of distance-one connections.

[6] Further details on the data source can be found in Section 4.1 below; for a description of the methodology and codebook, see Gallup (2013).

2.1. Intending migrants

The Gallup World Polls include two related questions on the intention to migrate, asked in all countries between 2007 and 2011: (*i*) "Ideally, if you had the opportunity, would you like to move to another country, or would you prefer to continue living in this country?", and (*ii*) "To which country would you like to move?" for the individuals who provide a positive answer to question (*i*). We refer to the individuals who express their intention to leave their country of residence as intending migrants.[7]



Notes: The figure plots the percentage of natives aged 15 to 49 intending to migrate from each country against the logarithm of real GDP per capita in 2010; data from the Gallup World Polls are pooled across different waves of the survey, and sampling weights are used; the surface of each circle is proportional to the size of the native population residing in each country.
Source: Authors' elaboration on Gallup World Polls and World Bank (2015a,b).

The average of the share of intending migrants, weighted by the size of the native resident population, stands at 21.1 percent.[8] The ten countries with the highest shares of intending

[7] The way in which this kind of hypothetical questions is interpreted might vary across countries, as observed by Clemens and Pritchett (2016), which is why we only use within-country variation in the estimation.

[8] Country-specific figures are aggregated using weights corresponding to the native population in each country in 2010, computed from World Bank (2015a,b), i.e., the size of the resident population minus the total number of foreign-born residents. Ideally, we would have used figures for the population aged 15 to 49, but these are not available neither for the resident population nor for the immigrant stocks. World Bank (2015a) does not provide an estimate of the total foreign-born population in Taiwan and in the Occupied Palestinian Territories, which we thus set to zero.

migrants among natives are either Sub-Saharan African or Latin American and Caribbean countries, with the Dominican Republic (65.9 percent) recording the largest share, followed by Sierra Leone (63.5), Haiti (62.8) and Guyana (62.1). Four out of the ten countries with the lowest shares of intending migrants are Gulf countries, namely Bahrain (2.6 percent), United Arab Emirates (4.5), Saudi Arabia (4.7) and Qatar (6.9).[9] The share of natives that intend to migrate declines with income per capita, as shown in Figure 1, with the bivariate correlation between the two variables standing at -0.265.

TABLE 1 • DISTRIBUTION OF INTENDING MIGRANTS BY DESTINATION COUNTRY

Share of intending migrants (percent)						
Destination	**World**	**Africa**	**America**	**Asia**	**Europe**	**Oceania**
United States	29.33	24.65	25.98	33.34	13.99	22.94
United Kingdom	7.94	10.55	8.73	6.86	9.87	22.11
Canada	6.48	5.49	9.07	5.98	7.29	14.23
France	5.66	10.46	6.46	4.24	6.81	4.78
Australia	4.40	0.79	2.63	5.31	6.07	6.57
Saudi Arabia	4.38	6.83	0.00	5.38	0.24	0.36
Japan	4.24	1.12	3.53	5.60	0.75	2.16
Germany	3.78	3.45	4.24	2.65	11.25	0.85
United Arab Emirates	2.94	2.32	0.01	4.08	0.46	0.86
Spain	2.89	2.29	12.09	0.29	8.17	1.26
South Korea	2.81	0.01	0.03	4.44	0.01	0.00
Singapore	2.76	0.01	0.00	4.35	0.08	1.49
Italy	2.63	3.61	5.15	1.54	4.89	2.47
Switzerland	1.49	0.47	1.24	1.56	2.98	0.00
Malaysia	1.37	0.16	0.00	2.13	0.07	0.12
Russia	1.36	0.28	0.22	1.77	1.85	0.51
China	0.82	1.02	1.34	0.74	0.26	0.75
Sweden	0.75	0.42	0.44	0.60	2.69	1.05
South Africa	0.73	4.95	0.23	0.08	0.17	1.70
New Zealand	0.73	0.07	0.10	0.83	1.79	4.60
Total top-20	87.47	78.96	81.49	91.77	79.67	88.81

Note: Share of intending migrations aged 15 to 49 across the top-20 countries of destination (defined at the world level), for the whole world and for each continent; data are pooled across countries and waves of the survey, and sampling weights are used to compute the distribution.
Source: Authors' elaboration on Gallup World Polls.

[9] India (6.7 percent), Thailand (9.4), Indonesia (10.7), China (11.1), Laos (11.4) and Malaysia (11.7) are the other countries with the lowest shares of intended migrants.

Table 1 reports the distribution of intending migrants across the top-20 countries of destination.[10] The natives aged 15 to 49 in our sample intend to migrate towards 185 different countries in the world, with a (highly) uneven distribution of intending migrants across (intended) destinations. Specifically, 29.3 percent of the individuals in our sample intend to migrate to the United States, followed by the United Kingdom (7.9), Canada (6.5), France (5.7) and Australia (4.8), with the first five (intended) destinations totaling 53.8 percent of the preferences of intending migrants. The top-20 intended destinations are chosen by around 87.5 percent of all intending migrants, while the total share of the 95 countries at the bottom of the list stands at just 1.0 percent. The (pooled) distribution of intending migrants across countries is closely and positively correlated with the distribution of actual migrant stocks, but it is more concentrated than the latter.[11] Table 1 also reveals the existence of relevant variations across continents in the distribution of intending migrants across destinations, although the top-20 destinations, defined at the world level, account for no less than 79.0 percent of migration intentions in each continent.

A reasonable concern might be that the answers to the hypothetical questions on migration intentions asked by Gallup are not informative about actual migration decisions. The OECD International Migration Database provides us with yearly data about the size of actual bilateral gross bilateral migration flows for 34 of the 185 destination countries mentioned as preferred destinations by the respondents to the Gallup World Polls.[12] Econometric analyses, presented in the Appendix A.1, reveal that bilateral migration intentions do contain relevant information about the size of actual bilateral migration flows.

2.2. Distance-one connections in the intended destinations

The questionnaire of the Gallup World Polls also includes the following question: (*iii*) "Do you have relatives or friends who are living in another country whom you can count on to help you when you need them, or not?". For the individuals who answer affirmatively to this question, the data provide (*iv*) information on up to three countries of residence of these relatives or friends.[13] Thus, questions (*iii*) and (*iv*) give us information about up to three countries in which each individual is directly connected to someone who could provide help to him or her.[14] 58 per-

[10] The respondents in each of the 147 countries in our sample differ with respect to the number of countries they intend to move to; on average, respondents in each country report 33.6 intended destinations, ranging from six for Trinidad and Tobago to 78 for Chad (see Table 2).

[11] The first ve intended destinations, which account for 53.8 percent of all intending migrants, hosted 35.9 percent of the actual migrants from the origin countries in our sample in 2010 according to World Bank (2015a).

[12] These 33 countries represent the preferred destination for 76.8 percent of the our sample of natives aged 15 to 49 who intend to migrate.

[13] The questionnaire also includes the following question: "Have any members of your household gone to live in a foreign country permanently or temporarily in the past ve years?", with information on the country of residence for those who provide an affirmative answer, but only for 287 out of 419 surveys; we do not employ this question in the analysis to avoid a substantial reduction in the sample size.

[14] Notice that questions (*iii*) and (*iv*) are asked in the Gallup World Polls before enquiring about the intentions to migrate, so that this dismisses the concern that respondents might be more likely to report a distance-one connection in the destination they intend to move to.

cent of the individuals who provide an affirmative answer to question (*iii*) report a distance-one connection in just one country, and 24 percent of them in two countries. This implies that for 82 percent of the respondents the limit of three countries in question (*iv*) is certainly not binding, so that we observe in the data all the countries in which they have a distance-one connection with relatives or friends, while the limit might be binding for (a part of) the 18 percent the respondents that report three countries. Thus, the Gallup World Polls give us information about the foreign countries in which each individual has at least one distance-one connection.

Notice that a respondent might have more than one distance-one connection in each of the countries that he or she reports, and that the distance-one connections might refer to individuals who are *not* born in the same country as the respondent. Keeping these two caveats in mind, it is interesting to compare the origin-specific distribution of the distance-one connections from the Gallup World Polls, conducted around the year 2010, with the actual distribution of its migrants across destinations in 2010 from World Bank 2015. For each country *j*, we compute the Spearman's rank correlation coefficient between the distributions of distance-one connections and actual migrants. This coefficient is always positive, and significantly so for 142 out of 144 countries,[15] and its (weighted) average stands at 0.519, with a standard deviation of 0.099.[16] The high value of the Spearman's rank correlation coefficient is reassuring with respect to the fact that the data coming out of the Gallup World Polls match well with the distribution of actual migrants across destinations.

3. The location-decision problem of intending migrants

Consider an individual i residing in country j, who has to select her preferred location from a choice set D. The utility that this individual would obtain from locating in country $k \in D$ is given by:

$$U_{ijk} = V_{ijk} + \varepsilon_{ijk} \tag{1}$$

where $V_{ijk} \equiv \mathrm{x}_{ijk}'\beta_{jk}$ represents the deterministic component of utility, net of moving costs, and ε_{ijk} is a stochastic term. If ε_{ijk} follows an independently and identically distributed Extreme Value Type-1 distribution, with $F(x) = e^{-e^{-x}}$, then the probability that country k represents the utility-maximizing choice is given by (McFadden 1974):

$$p_{ijk} \equiv \text{Prob}\left(U_{ijk} > U_{ijl} \ , \forall l \in D/\{k\}\right) = \frac{e^{\mathbf{x}_{ijk}'\boldsymbol{\beta}_{jk}}}{\sum_{l \in D} e^{\mathbf{x}_{ijl}'\boldsymbol{\beta}_{jl}}} \tag{2}$$

The separate estimation of a conditional logit model for each origin j allows us to recover the vectors of parameters β_{jk}. We model the deterministic component of utility as depending on a dummy variable d_{ijk} that signals whether the j-born individual i has a distance-one connection to destination k, and we denote by $\beta_{1jk} = \beta_{1j}$, $\forall k \in D$, the parameter associated to d_{ijk}.

[15] We do not have data on bilateral migrant stocks for the Occupied Palestinian Territories, Serbia and Taiwan from World Bank (2015a); the countries for which the Spearman's rank correlation coefficient is not significantly different from zero at the 1 percent confidence level are Bahrain (p-value 0.096) and Namibia (0.025).

[16] Similar evidence is obtained when relying on the Pearson's correlation coefficient.

The choice set over which we estimated (2) does not include the origin j itself, because the variable d_{ijk} cannot be properly defined when $k = j$, so that our estimation is restricted to the sub-sample of individuals stating an intention to migrate. Notice that the estimation on the choice set $D_j \equiv D / \{j\}$ entails that our estimation is consistent with the distributional assumptions introduced by Bertoli *et al.* 2013 and Ortega and Peri 2013, who allow for a common variance component of the stochastic term ε_{ijk} across all countries but the origin, which reflects unobserved individual heterogeneity in the preferences for migration, as this component does *not* influence the choice of the preferred option in D_j.[17]

The estimation of (2) rests on the independence of irrelevant alternatives property within the choice set D_j, which implies that the relative probability of choosing between two alternative options in D_j depends exclusively on the attractiveness of these two options, i.e., $\ln(p_{ijk}) - \ln(p_{ijh}) = V_{ijk} - V_{ijh}$, and it is independent from the presence of other alternatives in the choice set D_j.[18] An implication of this property is that the estimated coefficients should be stable when the choice set D_j is modified, as otherwise the relative choice probabilities would be altered. We thus re-estimate (2) on a series of restricted choice sets R_j^n that are obtained by dropping sets of destinations from D_j, comparing the estimated coefficient $\widehat{\beta}_{1j}^{R_j^n}$ obtained on the subsample $R_j^n \subset D_j$ with the point estimate $\widehat{\beta}_{1j}$ obtained from the estimation on the entire choice set D_j.[19] More specifically, for each country j we compute the share of the estimations conducted on the restricted samples R_j^n for which we do not reject the null hypothesis that $\widehat{\beta}_{1j}^{R_j^n} = \widehat{\beta}_{1j}$.[20]

4. Descriptive statistics

The Gallup World Polls cover the entire civilian, non-institutionalized population aged 15 years and above, with a sample of around 1,000 individuals in each wave of the survey. As discussed in Section 2 above, we restrict our sample to natives aged 15 to 49 who intend to migrate abroad.[21] The number of individuals included in the sample for each of the 147 countries depends on the number of waves of the Gallup World Polls conducted between 2007 and 2011, the share of foreign-born individuals residing in each country, and the share of intending migrants

[17] "The allocation of actual migrants by distance migrated should be relatively free of the influence of psychic costs, although the percentage of all persons who become migrants is not." (Sjaastad 1962).

[18] We should recall here that the independence of irrelevant alternatives is a property of the specification of the model that is estimated, rather than an inherent feature of the choice situation, and it depends on the extent to which observables allow capturing heterogeneity across individuals; Bertoli and Fernández-Huertas Moraga (2013, 2015) provide evidence that this property is violated in specifications estimated on aggregate data that assume that the deterministic component of utility is *not* individual-specific, while we relax this assumption in (2).

[19] See, for instance, Head *et al.* 1995 and Grogger and Hanson 2011.

[20] See Section 5.2 for more details.

[21] Foreign-born individuals are likely to have some unobserved characteristics, such as the proficiency in their mother tongue, that could be correlated both with the geographical distribution of their distance-one connections, and with the choice of their intended destination; 28.1 percent of the foreign-born intending migrants report their country of birth as their preferred destination, and 42.8 percent of them have a distance-one connection there.

in each country. Table 2 reports the number of waves of the Gallup World Polls for each country, together with the number of intending migrants among the natives aged 15 to 49 and the number of intended destinations. The total sample size is 86,875 intending migrants, which corresponds to an average of 591 per country, with the sample size varying between 29 (Bahrain) and 2,006 (Senegal).

38.0 percent of the 86,875 intending migrants in our sample have a distance-one connection in at least one foreign country, and 20.3 percent of the intending migrants have a distance-one connection in the destination they intend to move to.

Table 2 • Sample size and number of intended destinations

Country	Waves	Obs.	Dest.	Country	Waves	Obs.	Dest.
Algeria	2	279	22	Tunisia	3	517	33
Angola	1	189	23	Uganda	3	1310	50
Benin	1	125	28	Zambia	3	746	47
Botswana	2	586	39	Zimbabwe	3	1349	51
Burkina Faso	2	646	39	Argentina	3	458	33
Burundi	2	258	26	Belize	1	113	21
Cameroon	4	1858	59	Bolivia	4	998	32
Central African Republic	1	464	36	Brazil	2	320	30
Chad	4	999	78	Canada	3	198	41
Comoros	2	539	33	Chile	3	759	38
Congo (Kinshasa)	1	377	32	Colombia	4	1173	33
Congo Brazzaville	1	426	32	Costa Rica	3	596	31
Djibouti	3	589	39	Dominican Republic	4	1740	32
Egypt	2	315	24	Ecuador	3	521	28
Ghana	3	1432	44	El Salvador	4	1545	33
Guinea	1	366	28	Guatemala	4	979	31
Ivory Coast	1	274	24	Guyana	1	216	19
Kenya	3	1473	58	Haiti	2	429	34
Liberia	3	1579	46	Honduras	4	1426	30
Libya	1	209	16	Mexico	3	530	36
Madagascar	1	184	16	Nicaragua	4	1546	28
Malawi	1	370	23	Panama	3	530	30
Mali	3	850	46	Paraguay	2	206	17
Mauritania	4	776	46	Peru	4	1420	39
Morocco	2	408	20	Trinidad and Tobago	1	65	6
Mozambique	1	232	22	United States	2	185	31
Namibia	1	157	26	Uruguay	4	365	28
Niger	4	850	45	Venezuela	3	296	30
Nigeria	4	1912	55	Afghanistan	4	1030	41
Rwanda	2	227	29	Armenia	4	931	33
Senegal	4	2006	42	Azerbaijan	4	729	32
Sierra Leone	2	1104	36	Bahrain	2	29	12
Somalia	2	668	35	Bangladesh	4	1230	45
South Africa	4	666	46	Cambodia	4	1278	28

(continues)

(follows)

Sudan	2	489	41	China	3	1072	37
Tanzania	3	985	59	Georgia	4	725	34
Togo	1	229	27	Hong Kong	2	225	26
India	4	1052	31	Bulgaria	2	235	24
Indonesia	4	315	24	Croatia	4	281	25
Iran	2	512	34	Cyprus	2	230	28
Iraq	2	274	26	Czech Republic	3	264	34
Israel	4	419	33	Denmark	4	376	46
Japan	7	634	44	Estonia	3	373	29
Jordan	3	498	39	Finland	2	221	42
Kazakhstan	4	495	32	France	3	367	51
Kyrgyzstan	4	861	35	Germany	4	554	54
Laos	2	170	18	Greece	3	317	31
Lebanon	3	529	42	Hungary	3	448	32
Malaysia	4	342	30	Iceland	1	85	14
Mongolia	2	722	28	Ireland	3	293	23
Nepal	4	666	35	Italy	3	464	39
Occupied Palestinian Territory	3	427	33	Latvia	3	337	31
Pakistan	5	493	34	Lithuania	4	670	32
Philippines	4	1011	39	Luxembourg	2	179	29
Qatar	1	39	20	Macedonia	4	742	41
Russia	5	1435	57	Malta	2	286	26
Saudi Arabia	3	103	26	Moldova	4	1159	39
Singapore	5	533	30	Netherlands	2	206	33
South Korea	4	941	39	Norway	1	95	27
Sri Lanka	4	723	34	Poland	4	482	39
Syria	3	456	43	Portugal	3	361	35
Taiwan	2	486	33	Romania	3	480	31
Tajikistan	4	635	24	Serbia and Montenegro	4	1949	51
Thailand	3	204	31	Slovakia	1	209	21
Turkmenistan	1	169	20	Slovenia	2	204	31
United Arab Emirates	2	37	14	Spain	3	302	35
Uzbekistan	3	431	24	Sweden	3	401	44
Vietnam	2	292	20	Switzerland	1	56	25
Yemen	2	441	25	Turkey	3	393	51
Albania	4	974	26	Ukraine	4	692	42
Austria	3	205	35	United Kingdom	4	677	54
Belarus	4	693	42	Australia	2	204	29
Belgium	3	285	39	New Zealand	2	221	27
Bosnia and Herzegovina	4	687	35				

Notes: We report the number of waves of Gallup World Polls conducted in each country between 2007 and 2011, the number of natives aged 15 to 49 who intend to migrate and the number of intended destinations.
Source: Authors' elaboration on Gallup World Polls.

5. Estimation

The specification of the conditional logit model that we bring to the data includes: (*i*) a dummy variable d_{ijk} that signals whether the individual i has a distance-one connection in destination k; (*ii*) dyadic dummies $\mathbf{d}_{jk}$ that absorb the effect of all time-invariant dyadic (such as distance or linguistic proximity), origin or destination-specific variables, (*iii*) a vector z_{ij} of individual characteristics, including sex, four age cohorts,[22] and a dummy that takes the value one for individuals who completed at least nine years of education.[23] Importantly, notice that the inclusion of dyadic dummies $\mathbf{d}_{jk}$ also controls for the influence exerted by the size of the diaspora of j-born individuals in destination k on the choice of the (intended) destination, as this variable mostly evolves slowly over time, if this enters additively in the function that describes the deterministic component of location-specific utility V_{ijk} in (1).[24] The empirical specification is thus consistent with the econometric evidence provided with aggregate data by Beine *et al.* 2011 on the role of the size of the bilateral diaspora in shaping actual migration flows.[25] The conditional logit model is estimated separately for each of the 147 countries in our sample. Letting $N_j \equiv \#D_j$, the estimation of the conditional logit model requires estimating one coefficient of the alternative-specific variable d_{ijk} plus six times N_j–1 coefficients for the individual-specific variables and the destination-specific intercepts, i.e., a total of $1+6(N_j-1)$ coefficients. The standard errors for the estimated coefficients are obtained through bootstrapping (200 replications with replacement).

5.1. Benchmark specification

We focus our attention on the estimated coefficients $\widehat{\beta}_{1j}$, with j = 1, ..., 147, for our variable of interest d_{ijk}.[26] Figure 2 plots the estimated coefficient for distance-one connections for each country against the corresponding z-score. The estimated coefficients are always positive (ranging between 0.28 an 4.49), and significantly different from zero for 130 out of 147 countries, and the z-score falls short of the value that allows rejecting the null hypothesis at the 1 percent confidence level for countries that (mostly) have a very limited sample size, as Figure 2 reveals.

[22] Specifically, 15 to 19, 20 to 29, 30 to 39 and 40 to 49 years.

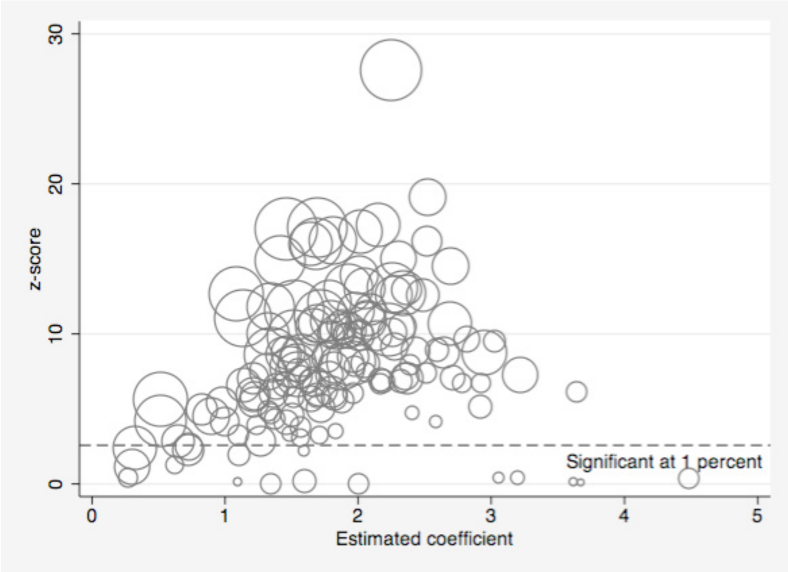
[23] The Gallup World Polls allow to distinguish three levels of education: up to eight years of schooling, from nine to 15 years, i.e., up to three years of post-secondary education, and completed tertiary education; our results are robust when including a dummy for each of the three levels, or when pooling together the two lowest levels education.

[24] We also present specifications where time-varying dyadic dummies, i.e., d_{jkt}, thus controlling also for variations over time in the size of the diaspora.

[25] Our specification is actually more general, as it does not require the diaspora to be defined on the basis of the country of birth; for instance, our specification can allow for the attractiveness of the United States for potential Ecuadorian migrants to depend on the size of the diaspora of all Spanish-speaking Latin American migrants residing in the United States.

[26] The minimal size N_j of the choice set for the countries in our sample is 14 (for Trinidad and Tobago), and it is thus unfeasible to report the $1+6(N_j-1) \geq 79$ estimated coefficients for each country.

FIGURE 2 • ESTIMATED COEFFICIENT AND Z-SCORE FOR DISTANCE-ONE CONNECTIONS



Notes: The figure plots country-specific point estimates for the coefficient of distance-one connections from the conditional logit and the corresponding *z*-score, (see also Table A.2 in the Appendix); the surface of each circle is proportional to the sample size for each country. Source: Authors' elaboration on Gallup World Polls and World Bank 2015a,b.

Figure plots the values of the estimated coefficients in a world map, and it reveals that there is no clear geographical pattern in the values of the estimates for the coefficient of distance-one connections.[27]

The average $\overline{\widehat{\beta}}_1$ of the estimated coefficients stands at 1.850, with a standard deviation of 0.689. This entails that the relative odds of intending to migrate to destination k over any other foreign destination for an individual with a distance-one connection in country k is around six to eight times larger than in the absence of a distance-one connection in k.[28]

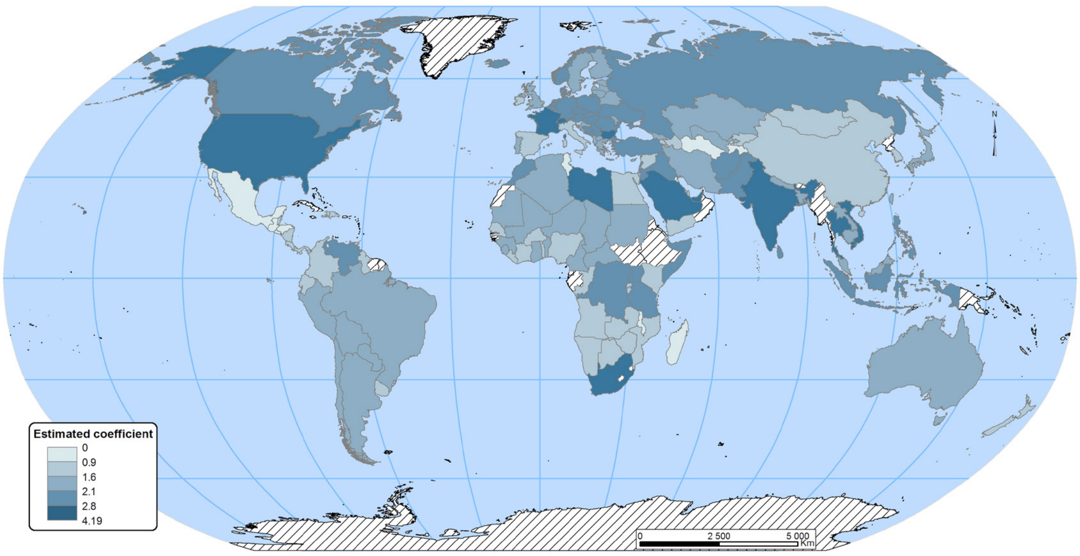
What can we say about the size of the estimated coefficient for distance-one connections? We cannot provide a direct comparison of our estimates with the effects of traditional determinants of (actual) migration decisions as the specification that we bring to the data controls for but does not provide an estimate for the effects of determinants of the attractiveness of a destination, such as its distance from the origin or the size of the diaspora, that do *not* vary across individuals. Still, the attractiveness of the various options in the choice set can be inferred from the estimated coefficients of the dyadic dummies $\mathbf{d}_{jk}$,

[27] Similar results are obtained when we estimate the model separately for men and women, or by level of education, or when we drop the individuals that report having friends and relatives they can count on in three distinct countries, as our variable of interest is probably measured with error as they might have distance-one connections in other countries, which would go unrecorded in the Gallup World Polls (see Section 2.2); the results are available from the authors upon request.

[28] We have that $e^{\overline{\widehat{\beta}}_1} \simeq 6.360$, while the average of the exponentiated values of the estimated coefficients stands at 8.395.

which reflect the differences in the deterministic component of location-specific utility,[29] and are thus directly comparable to $\widehat{\beta}_{1j}$. Given the distributional assumptions that we have introduced, the origin-specific distribution of the estimated values of the coefficients for the dummies $\mathbf{d}_{jk}$ is closely related to the distribution of observed choice probabilities, as the average of the individual-specific utility U_{ijk}, conditional upon k being the utility-maximizing alternative, is invariant with k (see de Palma and Kilani, 2007).[30] The distribution of migration intentions is very concentrated in a few destinations (see Section 2.1), and this, in turn, entails that the origin-specific distribution of the estimated coefficients for the dummies $\mathbf{d}_{jk}$ is very dispersed. Thus, $\widehat{\beta}_{1j}$ stands, on average, at 4.6 percent of the standard deviation of the distribution of the estimated coefficients for the dummies $\mathbf{d}_{jk}$, so that distance-one connections are *un*able to turn an otherwise unattractive destination into the preferred option for an intending migrant. Still, they do tilt the balance among countries that have a similar attractiveness, as main destinations do.

FIGURE 3 • ESTIMATED COEFFICIENTS FOR DISTANCE-ONE CONNECTIONS



Notes: The figure reports the estimates from the conditional logit (see Table A.2 in the Appendix). Source: Authors' elaboration on Gallup World Polls.

[29] More precisely, this is true for a woman aged 15 to 19 with no more than eight years of completed education; the difference in the deterministic component of utility for the respondents with other characteristics also depends on the destination-specific coefficients of the vector of individual-specific regressors $\mathbf{z}_{ij}$.

[30] $Uijk$ depends on the deterministic component V_{ijk} and on the stochastic component ε_{ijk}; if $V_{ijk} > V_{ijl}$, then destination k will represent the preferred option for a larger share of j–born intending migrants, and the average value of ε_{ijk} for them will be lower than the corresponding average value of ε_{ijl} for the individuals who intend to move to l, and this differential exactly offsets the difference between V_{ijk} and V_{ijl}, so that $E(U_{ijk}|U_{ijk} > U_{ijh}\ , \forall h \in D/\{k\}) = E(U_{ijl}|U_{ijl} > U_{ijh}\ , \forall h \in D/\{k\})$.

Our estimation approach is based on the assumption that the vector $\mathbf{x}_{ijk}$ is able to mop up all sources of correlation in utility U_{ijk} across the various options in the choice set. A violation of this identifying assumption could result in a bias in the estimate of β_{1j}. More specifically, an unobserved individual characteristic u_{ijk} that is positively correlated both with the dummy variable d_{ijk} that signals whether the j-born individual i has a distance-one connection in k and that contributes to increase the attractiveness of destination k would induce an upward bias in our estimate of β_{1j}, and it could introduce a correlation in utility across destinations. For instance, imagine that an intending migrant born in Argentina is of Italian origins: she is more likely to have a distance-one connection in Italy than other Argentine-born intending migrants, and she also faces lower legal barriers for migration to Italy (and to other EU member states), as any foreign-born individual of proven Italian descent can obtain the Italian citizenship (Law No. 91, February 5, 1992). The resulting omitted variable bias could produce a positive and significant estimate for β_{1j} even in the absence of any causal effect, and it would result in a violation of the independence of irrelevant alternatives property. We thus check whether the specification that we bring to the data satisfies the IIA property, and we then explicitly deal with threats to our identification strategy that can be due to a number of plausible *un*observed factors.

5.2. Testing for the IIA property

The estimation of the conditional logit model rests on the property of the independence of irrelevant alternatives, as discussed in Section 3 above. We test whether the estimate of is $\widehat{\beta}_{1j}$ stable when we re-estimate the model on a restricted choice set. Specifically, for each estimation on a restricted sample R_j^n, we see whether the estimated coefficient $\widehat{\beta}_{1j}^{R_j^n}$ falls within the 95 percent confidence interval of $\widehat{\beta}_{1j}$, i.e., $\widehat{\beta}_{1j}^{R_j^n} \simeq \widehat{\beta}_{1j}$; we then compute the share of the estimations for which this is actually the case.[31] We follow two distinct approaches to define the restricted samples R_j over which the conditional logit is estimated: (*i*) we drop one (intended) destination at a time, as in Grogger and Hanson 2011, so that $n = 1, ..., N_j$; (*ii*) we sort the countries in the choice set D_j in ascending order of the number of intending migrants, and we drop larger sets of destinations starting from the one with the lowest number of intending migrants. The second approach is clearly more demanding, as the size of the restricted sample R_j gets progressively smaller.[32]

On average, 98.5 percent of the specifications defined on the basis of the approach described at point (*i*) produce an estimated coefficient for distance-one connections which belongs to the 95 percent confidence interval of $\widehat{\beta}_{1j}$. When we follow the more demanding approach described in (*ii*) which induces major reductions in the dimension of the choice set and in the sample size, we find that 90.9 percent of the specifications produce an estimated coefficient for d_{ijk} that lies in the confidence interval of the one obtained from our

31 This test requires estimating the conditional logit model more than 12,000 times, which is why we do not bootstrap standard errors for the specifications estimated on the restricted samples.

32 The number of replications in this second approach is not higher than N_j-2, as the conditional logit might fail to converge when just a few destinations are included in R_j^n.

benchmark specification. Both approaches are thus reassuring about the appropriateness of the IIA property that characterizes the specification of the location-choice model that we have brought to the data.

5.3. Is our estimate just capturing correlated peer effects?

As discussed above, the estimated effect of distance-one connection might be due to unobserved variables that are correlated both with our variable of interest and with location-specific utility. We follow three distinct but complementary approaches to mitigate the concerns that our evidence about the key role played by distance-one connections in determining the preferred intended destinations is just reflecting correlated peer effects. Specifically, (*i*) we add further individual-level variables to the vector $\mathbf{z}_{ij}$, and (*ii*) we re-estimate the model on a suitably defined set of destinations.[33]

5.3.1. Inclusion of additional controls

Our benchmark specification includes an origin-destination specific intercept of the deterministic component of utility V_{ijk}. As we pool the data from the Gallup World Polls across waves, one might be concerned that the attractiveness of destination k for j-born intending migrants might vary over time, and that these variations could be correlated with the likelihood of having a distance-one connection there. For instance, sustained economic growth in k could both attract more migrants from country j, thus increasing the number of non-migrants that have a distance-one connection in k, and it could increase the share of j-born intending migrants for which k represents the preferred destination. We re-estimate the conditional logit model allowing the origin-destination specific intercept to vary with each wave of the Gallup World Polls:[34] the correlation of the ensuing set of coefficients with those from our benchmark specification stands at 0.992.

We also include additional elements to the vector $\mathbf{z}_{ij}$ relying on information contained in the Gallup World Polls. Specifically, we separately add (detailed) dummies for the self-reported religion of each respondent,[35] and an asset index *à la* Dustmann and Okatenko (2014).[36] The first of the two extensions of our benchmark specification allows to dismiss the concern that religion might influence both individual preferences across destinations and the geographical

[33] All the results that are discussed but not reported are available from the authors upon request.

[34] We have more than one wave for 124 out of 147 countries (see Table 2).

[35] Information about religion is available for 142 out of 147 countries in our sample.

[36] Specifically, the asset index is the first principal component computed through an origin-specific polychoric principal component analysis on four of the seven questions used by Dustmann and Okatenko (2014) that are available for all countries in our sample from 2007 to 2011; the questions relate to (*i*) the ownership of a TV set, (*ii*) access to the Internet, to whether in the previous 12 months the respondent did not have enough money (*iii*) to buy food or (*iv*) to provide adequate shelter of housing to her family.

distribution of one's own distance-one connections.[37] The second extension deals with the concern related to a different form of homophily, as an individual is likely to be mostly connected with other individuals with a similar socio-economic condition, which could influence the set of destinations that an individual can afford to move to. Allowing location-specific utility to vary either across religious groups or with the household's socio-economic status, as proxied by the asset index, does *not* result in a significant reduction in the estimated values of $\widehat{\beta}_{1j}$, which remain closely correlated with those obtained in the benchmark specification.

5.3.2. Restrictions of the choice set

A different way to deal with the threats to identification posed by individual-level unobservables is through suitable restrictions of the choice set. For instance, one might be concerned that the (unobserved) proficiency in a foreign language influences both the expected returns from migration to the countries where this language is spoken, and the distribution of one's own distance-one connections. We thus restrict the choice set to destinations where English is (one of) the official language(s).[38] English is an official language in seven out of the top-20 intended destinations in Table 2; on average, 46.0 percent of the intending migrants report an English-speaking country as their preferred destination, and this figure is not lower than 30.0 percent for three out of four countries in our sample.[39] The *un*observed proficiency in English, which is potentially correlated with the likelihood of having a distance-one connection in an English-speaking country, cannot influence the choice of the intended destinations within the restricted choice set of English-speaking destinations. Once again, the results from our benchmark specification do not appear to be sensitive to this threat to identification: the estimated coefficients in the restricted choice set are *not* systematically lower than in the entire choice set, where the spurious correlation of d_{ijk} with unobserved proficiency in English could have imparted an upward bias in our estimate of β_{1j}.

The Gallup World Polls provide information on the country of birth of each respondent, so that we can restrict our sample to native-born only, as discussed in Section 4. Nevertheless, some of the natives could be of immigrant descent,[40] and these individuals might differ from the rest of the sample in similar unobserved dimensions as foreign-born respondents do. We thus rely on data from World Bank (2015a) to identify the ten countries with the largest stock of immigrants residing in country j in 2010, and we exclude

[37] For instance, a Muslim born in Egypt could be more likely to have distance-one connections in Gulf countries and to intend to migrate there, while a Coptic Christian born in the same country could be more likely to have distance-one connections in the United States and to state her intention to move to this destination.

[38] The size of the of restricted choice set varies from three (for Egypt, Libya, Qatar and Venezuela) to 25 (for Kenya).

[39] The corresponding figures are much lower for subsets of destinations that share another official language, such as Spanish, Arabic or Russian, which prevents the estimation on these restricted choice sets.

[40] Later waves of the Gallup World Polls allow identifying second-generation immigrants, but they do not contain information on distance-one connections.

these countries from the choice set of *j*-born intending migrants.[41] Following up on the example introduced in Section 5.1, this criterion ensures that we drop Italy from the choice set of Argentine-born intending migrants, as Italians are one of the largest immigrant groups in Argentina. This addresses the threat to identification posed by the fact that natives of immigrant descent might face lower moving costs – for legal, linguistic or cultural reasons – to the country of origin of their ancestors, where they are also likely to have a distance-one connection.

The main countries of intended migration can also be the countries of origin of the largest immigrant stocks for some countries in our sample, so that this criterion at times leads to a drastic reduction in the sample size that produces outliers in the estimation.[42]

This restriction in the choice set does *not* result in a systematic reduction in the estimated effect of distance-one connections, as the correlation of the point estimates with those from our benchmark specification stands at 0.391.[43]

6. Concluding remarks

This paper relies on individual-level data from the Gallup World Polls to provide econometric evidence on the relationship between an individual's direct connections to the migrant networks in different countries and her choice concerning the preferred country of destination. The data from the Gallup World Polls give us a much finer measure of migrant networks than those commonly employed in the literature, which allow us to get a deeper understanding of the way in which networks influence migration decisions.

Distance-one connections appear to be a key driver in the choice among competing destinations with a similar level of attractiveness. The estimated effect is small relative to the dispersion of the levels of attractiveness of the various countries which are implied by the identifying assumption that stated preferences among competing destinations reflect an utility-maximizing behavior. We present various robustness checks which allow to mitigate the concern that unobserved individual heterogeneity is driving the estimated effects of distance-one connections.

[41] We obtain similar results when relying on migrant stocks data for earlier decades from Özden *et al.* (2011), as the set of main origin countries tends to remain unchanged over time.

[42] For instance, eight of the ten main countries of origin of the immigrants in Guyana are also among the top ten countries of intended migration according to the Gallup World Polls, so that less than 8 percent of its intending migrants belong to the restricted sample.

[43] As recalled above, World Bank (2015a) does not provide information on bilateral immigrant stocks for the Occupied Palestinian Territories, Serbia and Taiwan; estimates for five countries (Belize, Guyana, Iceland, Switzerland and Trinidad and Tobago) with outlying values of the estimated coefficients have been excluded when computing the correlation.

7. References

Beine M., Docquier F. and Özden Ç. (2011), *Diasporas*, «Journal of Development Economics», 95, 1, pp. 30-41

– (2015), *Dissecting network externalities in international migration*, «Journal of Demographic Economics», 81, 4, pp. 379-408

Beine M. and Salomone S. (2013), Network Effects in International Migration: Education versus Gender, «The Scandinavian Journal of Economics», 115, 2, pp. 354-380

Bertoli S. (2010), *Networks, sorting and self-selection of Ecuadorian migrants*, «Annales d'Economie et de Statistique», 97/98, pp. 261-288

Bertoli S. and Fernández-Huertas Moraga J. (2013), *Multilateral Resistance to Migration*, «Journal of Development Economics», 102, pp. 79-100

– (2015),*The size of the cliff at the border*, «Regional Science and Urban Economics», 51, pp. 1-6.

Bertoli S., Fernández-Huertas Moraga J. and Ortega F. (2013), *Crossing the border: Self-selection, earnings and individual migration decisions*, «Journal of Development Economics», 101, 1, pp. 75-91

Carrington W.J., Detragiache E. and Vishwanath T. (1996), *Migration with endogenous moving costs*, «American Economic Review», 86, 4, pp. 909-930

Chort I. (2014), *Mexican migrants to the US: What do unrealized migration intentions tell us about gender inequalities?*, «World Development», 59, pp. 535-552

Clemens M. and Pritchett L. (2016), *The New Economic Case for Migration Restrictions: An Assessment*, Working Paper No. 423, Washington, Center for Global Development

Comola M. and Mendola M. (2015), Formation of Migrant Networks, «The Scandinavian Journal of Economics», 117, 2, pp. 592-618

Creighton M. (2013), *The Role of Aspirations in Domestic and International Migration*, «Social Science Journal», 50, 1, pp. 79-88

De Palma A. and Kilani K. (2007), *Invariance of conditional maximum utility*, «Journal of Economic Theory», 132, 1, pp. 137-146

Delogu M., Docquier F. and Machado J. (2015), *Globalizing Labor and the World Economy: The Role of Human Capital*, mimeo, Université Catholique de Louvain

Docquier F., Machado J. and Sekkat K. (2015), *Efficiency gains from liberalizing labor mobility*, «The Scandinavian Journal of Economics», 117, 2, pp. 303-346

Docquier F., Peri G. and Ruyssen I. (2014), *The cross-country determinants of potential and actual migration*, «International Migration Review», 48, S37-S99

Dustmann C. and Okatenko A. (2014), Out-migration, wealth constraints, and the quality of local amenities, «Journal of Development Economics», 110, pp. 52-63

Esipova N., Ray J. and Pugliese A. (2011), *Gallup World Poll: The Many Faces of Global Migration*, IOM Migration Research Series 43, Geneva, International Organization for Migration

Gallup (2013), *Worldwide Research Methodology and Codebook*, Washington, D.C.

Giulietti C., Wahba J. and Zenou Y. (2014), *Strong Versus Weak Ties in Migration*, Discussion Paper No. 10248, Lond, CEPR

Grogger J. and Hanson G.H. (2011), Income maximization and the selection and sorting of international migrants, «Journal of Development Economics», 95, 1, pp. 42-57

Guimaraes P., Figueiredo O. and Woodward D. (2003), *A tractable approach to the firm location decision problem*, «Review of Economics and Statistics», 85, 1, pp. 201-204

Head K., Ries J. and Swenson D. (1995), *Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States*, «Journal of International Economics», 38, 3-4, pp. 223-247

Jackson M.O. (2010), *Social and Economic Networks*, Princeton, Princeton University Press

Manchin M., Manchin R. and Orazbayev S. (2014), *Desire to Migrate Internationally and Locally and the Importance of Satisfaction with Amenities*, Paper presented at the FIW-wiiw Seminars in International Economics, April 10, 2014, Wien

Mayer T. and Zignago S. (2011), *Notes on CEPII's distances measures: The GeoDist database*, Working Paper No. 2011-25, Paris, CEPII

McFadden D. (1974), *Conditional logit analysis of qualitative choice behavior*, in P. Zarembka (ed.), *Frontiers in Econometrics*, New York, Academic Press, pp. 105-142

McKenzie D.J. and Rapoport H. (2010, *Self-selection patterns in Mexico-U.S. migration: The role of migration networks*, «Review of Economics and Statistics», 92, 4, pp. 811-821

Munshi K. (2003), *Networks in the modern economy: Mexican migrants in the U.S. labor market*, «Quarterly Journal of Economics», 118, 2, pp. 549-599

– (2016), *Community networks and migration*, in J. Bramoullé, A. Galeotti and B.W. Rogers (eds), *Oxford Handbook on the Economics of Networks*, Oxford, Oxford University Press, pp. 630-648

Munshi K. and Rosenzweig M. (2016), *Networks and misallocation: Insurance, migration, and the rural-urban wage gap*, «American Economic Review», 106, 1, pp. 46-98

Ortega F. and Peri G. (2013), The Role of Income and Immigration Policies in Attracting International Migrants, «Migration Studies», 1, 1, pp. 27-46

Özden Ç., Parsons C.R., Schiff M. and Walmsley T.L. (2011), *Where on earth is everybody? The evolution of global bilateral migration 1960-2000*, «World Bank Economic Review», 25, 1, pp. 12-56

Patel K. and Vella F. (2013), *Immigrant networks and their implications for occupational choice and wages*, «Review of Economics and », 95, 4, pp. 1249-1277

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

Sjaastad L.A. (1962), *The Costs and Returns of Human Migration*, «Journal of Political Economy», 70, 1, pp. 80-93

World Bank (2015a), *Bilateral migration matrix 2010*, Washington D.C., World Bank, http://go.worldbank.org/JITC7NYTT0 (accessed on October 10, 2015)

– (2015b), *World Development Indicators*, Washington D.C., World Bank

8. Appendix

8.1 Intentions to migrate and actual migration

The data from the Gallup World Polls can be aggregated to obtain the number of natives of country j intending to move to country k in each year in which the survey is conducted, which we denote as intention$_{jkt}$. The OECD International Migration Database provides us with information about the size of the actual gross bilateral migration flow from j to k by year, which we denote by flow$_{jkt}$, for 34 of the 185 destination countries mentioned as preferred destinations by the respondents to the Gallup World Polls. We can then test whether the number of intending migrants contains information about the size of actual bilateral migration flows once we control for a number of origin-specific, destination-specific or dyadic factors with a Poisson Pseudo Maximum Likelihood estimation. Specifically, we estimate the following regression:

$$\text{flow}_{jkt} = \exp\left[\alpha \ln(\text{intention}_{jkt}) + \beta' x_{jk} + \mathbf{d}_{jt} + \mathbf{d}_{kt} + \varepsilon_{jkt}\right]$$

where x_{jk} is a vector of dyadic controls including the logarithm of distance, and dummies for contiguity, common colonial history and a common language, and $\mathbf{d}_{jt}$ and $\mathbf{d}_{kt}$ represent origin-year and destination-year dummies respectively. We also estimate (A.1) collapsing the longitudinal dimension of the data,[44] and including the logarithm of the size of the bilateral migration stock as an additional element in $\mathbf{x}_{jk}$, following Beine *et al.* 2011.

Table reports the estimates of the various specifications of (A.1): the estimated elasticity of bilateral migration flows with respect to the number of bilateral intending migrants stands at 0.627-0.800 in the cross-sectional analysis, and at 0.409-0.540 when the longitudinal dimension of the data is used. The estimated elasticity is positive and highly statistically significant even in the fourth data column of Table A.1, where we control for the time-varying attractiveness of each destination and for the size of the diaspora. Similar results, reported in the last two data columns of Table A.1, are obtained when we exclude high-income origin countries from the sample, as natives of those countries could be better able to turn their intentions into actual migration episodes.

[44] The data are collapsed over the (dyad-specific) set of years for which the information on bilateral migration intentions from the Gallup World Polls is not missing.

Table A.1 • Migration intentions and actual migration flows to OECD destinations

Specification *Dependent variable*	*(1)* *$flow_{jk}$*	*(2)* *$flow_{jk}$*	*(3)* *$flow_{jkt}$*	*(4)* *$flow_{jkt}$*	*(5)* *low_{jkt}*	*(6)* *$flow_{jkt}$*
ln($intentions_{jkt}$)	0.800*** [0.048]	0.627*** [0.038]	0.540*** [0.028]	0.409*** [0.027]	0.444*** [0.032]	0.345*** [0.033]
ln($networks_{jk}$)		0.247*** [0.038]		0.242*** [0.022]		0.192*** [0.028]
ln($distance_{jk}$)	-0.588*** [0.066]	-0.401*** [0.060]	-0.712*** [0.045]	-0.496*** [0.049]	-1.031*** [0.056]	-0.816*** [0.055]
$Contiguity_{jk}$	0.585*** [0.167]	0.372** [0.148]	0.506*** [0.095]	0.314*** [0.086]	1.556*** [0.159]	1.081*** [0.154]
Common $language_{jk}$	0.318** [0.130]	0.371*** [0.119]	0.515*** [0.073]	0.529*** [0.068]	0.583*** [0.087]	0.650*** [0.091]
$Colony_{jk}$	0.308** [0.132]	0.033 [0.117]	0.348*** [0.056]	0.065 [0.061]	0.434*** [0.077]	0.109 [0.098]
Destination dummies	Yes	Yes	No	No	No	No
Destination-year dummies	No	No	Yes	Yes	Yes	Yes
Origin dummies	Yes	Yes	No	No	No	No
Origin-year dummies	No	No	Yes	Yes	Yes	Yes
Observations	2,512	2,512	4,534	4,534	2,872	2,872
Pseudo-R^2	0.854	0.890	0.878	0.907	0.939	0.948

Note: standard errors in brackets; *** signicant at the 1 percent level, ** signicant at the 5 percent level, * signicant at the 10 percent level; the dependent variable in specications (1)-(2) is obtained collapsing the variables for each origin-destination pair over time before taking the logarithmic transformation; specica-tions (5)-(6) exclude from the sample the origin countries that are classied as high-income countries by the World Bank.
Source: Authors' elaboration on Gallup World Polls, OECD International Migration Database, Mayer and Zignago (2011) and Özden *et al*. (2011).

8.2 Benchmark estimates

Table 2 • Estimated coefficients for distance-one connections

Country	obs.	coeff.	s.e.	Country	obs.	coeff.	s.e.
Algeria	279	1.606	8.425	Tunisia	517	0.725	0.331
Angola	189	1.246	0.324	Uganda	1310	2.261	0.173
Benin	125	1.842	0.527	Zambia	746	1.548	0.192
Botswana	586	1.140	0.174	Zimbabwe	1349	1.422	0.096
Burkina Faso	646	1.444	0.169	Argentina	458	1.773	0.212
Burundi	258	1.111	0.595	Belize	113	1.492	0.453
Cameroon	1858	1.695	0.099	Bolivia	998	1.644	0.103
Central African Republic	464	1.680	0.248	Brazil	320	1.886	0.345
Chad	999	1.789	0.148	Canada	198	2.788	0.420
Comoros	539	0.654	0.232	Chile	759	1.695	0.159
Congo (Kinshasa)	377	2.275	0.250	Colombia	1173	1.344	0.114
Congo Brazzaville	426	1.000	0.241	Costa Rica	596	1.215	0.206
Djibouti	589	1.474	0.169	Dominican Republic	1740	1.143	0.104
Egypt	315	1.469	0.361	Ecuador	521	1.213	0.173
Ghana	1432	1.722	0.155	El Salvador	1545	0.515	0.093
Guinea	366	1.832	0.321	Guatemala	979	0.329	0.137
Ivory Coast	274	1.514	0.333	Guyana	216	1.382	0.324
Kenya	1473	1.522	0.155	Haiti	429	1.374	0.229
Liberia	1579	1.352	0.157	Honduras	1426	0.524	0.126
Libya	209	4.489	11.802	Mexico	530	0.835	0.169
Madagascar	184	0.275	0.808	Nicaragua	1546	1.084	0.086
Malawi	370	0.729	0.294	Panama	530	0.985	0.185
Mali	850	1.799	0.166	Paraguay	206	1.981	0.254
Mauritania	776	2.079	0.192	Peru	1420	1.689	0.106
Morocco	408	2.262	0.225	Trinidad and Tobago	65	1.598	0.744
Mozambique	232	1.351	0.290	United States	185	2.930	0.435
Namibia	157	1.573	0.520	Uruguay	365	1.394	0.216
Niger	850	2.054	0.158	Venezuela	296	2.177	0.318
Nigeria	1912	1.527	0.133	Afghanistan	1030	2.344	0.184
Rwanda	227	1.575	0.422	Armenia	931	1.329	0.133
Senegal	2006	1.460	0.086	Azerbaijan	729	0.906	0.203
Sierra Leone	1104	1.876	0.240	Bahrain	29	3.676	114.676
Somalia	668	2.314	0.155	Bangladesh	1230	1.927	0.148
South Africa	666	3.229	0.448	Cambodia	1278	1.957	0.241
Sudan	489	1.882	0.179	China	1072	1.557	0.215
Tanzania	985	2.694	0.254	Georgia	725	1.677	0.162

Togo	229	1.108	0.349	Hong Kong	225	1.525	0.264
India	1052	2.957	0.338	Bulgaria	235	3.028	0.317
Indonesia	315	2.717	0.384	Croatia	281	1.646	0.240
Iran	512	1.972	0.233	Cyprus	230	1.899	0.191
Iraq	274	2.599	0.292	Czech Republic	264	2.318	0.338
Israel	411	2.054	0.256	Denmark	376	2.333	0.223
Japan	634	1.721	0.272	Estonia	373	1.790	0.247
Jordan	498	2.647	0.303	Finland	221	2.009	7.1*104
Kazakhstan	495	1.827	0.235	France	367	2.827	0.295
Kyrgyzstan	861	1.496	0.191	Germany	554	2.493	0.199
Laos	170	1.712	0.529	Greece	317	1.697	0.291
Lebanon	529	2.106	0.181	Hungary	448	2.148	0.219
Malaysia	342	1.654	0.294	Iceland	85	2.588	0.636
Mongolia	722	1.328	0.179	Ireland	293	1.554	0.208
Nepal	666	1.932	0.190	Italy	464	1.219	0.223
Occupied Palestinian Terr.	427	2.433	0.276	Latvia	337	1.938	0.196
Pakistan	493	2.369	0.340	Lithuania	670	1.854	0.183
Philippines	1011	2.156	0.125	Luxembourg	179	2.063	0.277
Qatar	39	1.099	17.822	Macedonia	742	2.008	0.144
Russia	1435	2.228	0.218	Malta	286	1.568	0.220
Saudi Arabia	103	3.203	8.453	Moldova	1159	1.809	0.111
Singapore	533	1.810	0.298	Netherlands	206	2.174	0.326
South Korea	941	1.586	0.186	Norway	95	2.407	0.508
Sri Lanka	723	2.701	0.187	Poland	482	2.368	0.184
Syria	456	1.273	0.456	Portugal	361	2.020	0.201
Taiwan	486	2.015	0.206	Romania	480	2.525	0.157
Tajikistan	635	0.301	0.260	Serbia / Montenegro	1949	2.255	0.082
Thailand	204	3.643	0.595	Slovakia	209	2.520	0.343
Turkmenistan	169	0.625	0.529	Slovenia	204	1.346	659.144
United Arab Emirates	37	3.625	26.769	Spain	302	1.458	0.222
Uzbekistan	431	1.727	0.346	Sweden	401	1.807	0.183
Vietnam	292	2.926	0.580	Switzerland	56	3.057	8.771
Yemen	441	1.225	0.211	Turkey	393	2.361	0.331
Albania	974	2.027	0.121	Ukraine	692	2.267	0.180
Austria	205	2.395	0.304	United Kingdom	677	2.076	0.189
Belarus	693	1.988	0.173	Australia	204	1.970	0.328
Belgium	285	2.196	0.323	New Zealand	221	1.328	0.276
Bosnia and Herzegovina	687	2.520	0.132				

Notes: standard errors obtained through bootstrapping with replacement, 200 replications.
Source: Authors' elaboration on Gallup World Polls.

Ainhoa Aparicio Fenoll
Zoë Kuehn

Education Policies and Migration across European Countries[1]

Abstract. This paper tests whether and how two education policies: (i) increasing the length of compulsory education and (ii) introducing foreign languages into compulsory school curricula, affect subsequent migration across European countries. We construct a novel data base that includes information on education reforms for thirty-one countries spanning four decades. Combining this data with information on recent migration flows by cohorts, we find that an additional year of compulsory education reduces the number of emigrants by almost 10 per cent. Increasing the length of compulsory education shifts educational attainment for a significant fraction of the population from low towards medium levels. Our findings are thus in line with the fact that in the majority of European countries medium educated individuals display lower emigration rates than low educated individuals. Introducing a foreign language into compulsory school curricula on the other hand, almost doubles the number of emigrants to the country where the language is spoken and increases the total number of emigrants by 20 per cent. Depending on the specific content of an education policy, "more education" can thus have opposite effects on migration.

Keywords. migration, compulsory schooling, foreign language proficiency, education

1. Introduction

More than fifty years after the signing of the first European treaty on labor mobility (Treaty of Rome: 1957), large differences in national unemployment rates across Europe persist. In 2014, in Spain and Greece unemployment was 24-26 per cent, while in contrast Germany with 5 per cent had one of the lowest unemployment rates (see Figure 1.1). According to the OECD, annual migration rates across European Union (EU) countries were around 0.3 per cent in 2010 while US state-to-state migration rates were 2.4 per cent[2]. As a response to the latest economic crisis, migration across European countries and in particular from Portugal, Italy, and Spain to

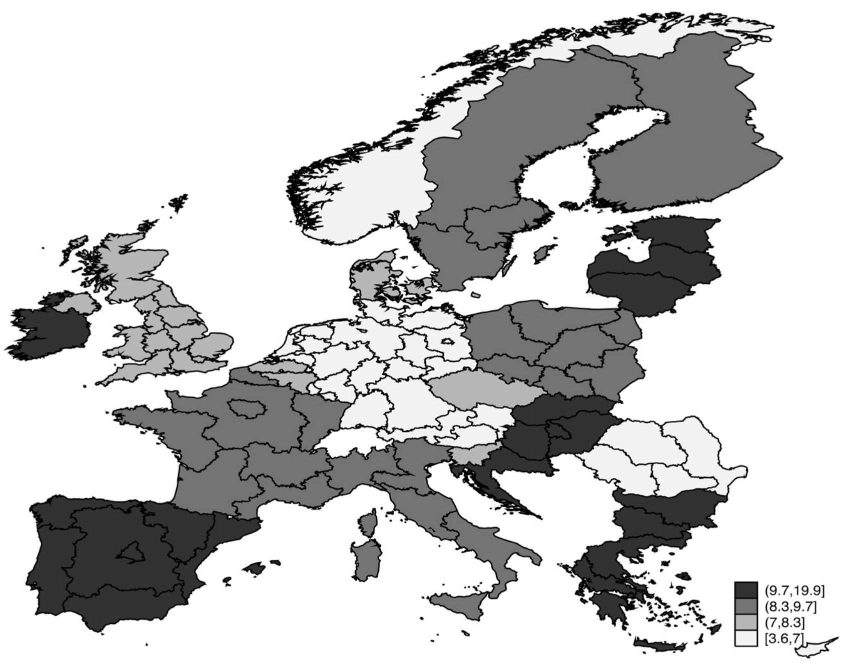
[1] We would like to thank Jesus Fernandez-Huertas Moraga, David McKenzie, Jennifer Graves, and seminar participants at Collegio Carlo Alberto, the 14th IZA/SOLE Transatlantic Meeting of Labor Economists, and the Workshop on Migration Barriers in Jena for their helpful comments and suggestions.

[2] Differences in US unemployment rates by state are much smaller. In 2014, they ranged from 2.8 per cent in North Dakota to 7.8 per cent in Washington D.C, Mississippi, and Nevada (Bureau of Labor Statistics). US migration rates are about twice as large as within-country migration rates in most European countries with the exception of Scandinavian countries and Great Britain, see Gakova and Dijkstra (1995) and Molloy, Smith and Wozniak (2011).

Germany has increased somewhat, see Jauer *et al.* (2014)[3]. However, overall labor mobility remains limited, unlikely to significantly reduce the observed differences in unemployment rates of 20 percentage points.

Language barriers seem to be an obvious explanation for the relatively low European labor mobility. The European Union consists of 28 countries and has 24 official working languages. Furthermore, these languages differ quite a bit. Linguists identify at least seven different language families among them: celtic, italic, germanic, baltic-slavik, greek, uralic, semitic; see Gray and Atkinson (2003) and Harding and Sokal (1988). In the US on the other hand, English is the only official language. Results in Bartz and Fuchs-Schündeln (2012) show that in Europe language barriers more than country borders hinder migration. Machin, Salvanes and Pelkonen (2012) bring forward an additional explanation. The authors suggest that lower educational attainment in Europe compared to the US leads to lower mobility[4].

FIGURE 1.1 • UNEMPLOYMENT RATES ACROSS EUROPE, 2014



Data: Eurostat

Education policies which increase schooling and improve foreign language proficiency affect educational attainment and language barriers, respectively. This suggests that education reforms – in particular of compulsory education which concerns all students – have

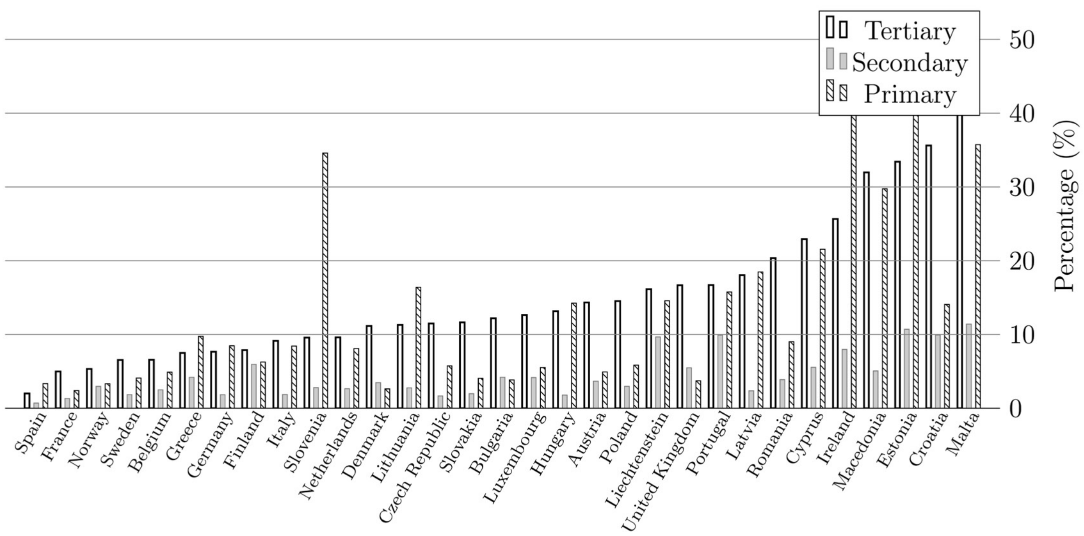
[3] EU law guarantees free labor mobility but countries can impose temporary restrictions for nationals of new member states. Prior to 2014, some EU member states required that Bulgarian and Romanian nationals obtained residence and work permits (see European Commission).

[4] Alternative explanations for the low European mobility focus on relatively high unemployment benefits (Antolin and Bover 1997) and stronger employment protection (Belot 2007) in European countries compared to the US.

the potential to affect migration. The current paper tests how across-country-and-time differences in such education policies affect recent European migration. In particular, we consider the following two policies that have been put into practice repeatedly by many countries worldwide: (i) increasing the length of compulsory education and (ii) introducing foreign languages into compulsory school curricula. Our findings show that additional years of compulsory education which shift the educational attainment for a significant fraction of the population from low towards medium levels decrease migration. While this result stands in contrast to most findings in literature on internal migration, it is in line with lower cross-country emigration rates of medium educated individuals compared to low educated individuals in the majority of European countries. On the other hand, we find that introducing a foreign language into compulsory school curricula increases migration to the country where the language is spoken.

Figure 1.2 displays emigration rates for 2010 by educational attainment for individuals age 25 and older in thirty-one European countries. With the exception of Denmark, Bulgaria, and the United Kingdom (UK), those with secondary educational attainment display lower emigration rates compared to primary and tertiary educated. Hence, in most European countries, the relationship between education and migration displays a u-shaped pattern. Moreover, in most European countries more than 40 per cent of individuals in the age groups most likely to migrate for job-related reasons (age 25-44) only completed their secondary education. In the US, on the other hand this is only the case for 32 per cent of individuals, see Figure A-1 of the Appendix. Hence, the low propensity to migrate of secondary educated individuals who make up an important fraction of the active population might explain the relatively low aggregate mobility in Europe.

Figure 1.2 • Emigration rates of individuals > 25 by educational attainment, 2010



Source: Brücker, Capuano, and Marfouk [2013]; primary: no schooling, primary and lower secondary; secondary: high-school leaving certicate or equivalent; tertiary: higher than high-school leaving certicate or equivalent.

But why would secondary educated individuals be less likely to migrate than primary and tertiary educated individuals? In the next section we present a model that is able to rationalize such a u-shaped relationship between education and migration. If transferring education across countries is costly, wages are increasing in education, returns to education are higher in the destination country than in the country of origin, and there exists a minimum wage paid independently of one's education in the destination country, then only low and high, but not medium educated, decide to migrate. We then introduce foreign language proficiency into the model. Foreign languages are necessary for transferring human capital across countries, and better language skills reduce the wage penalty for immigrants. We use this model to predict the impact of education policies on migration.

In particular, we foresee that when migration rates are u-shaped, an increase in years of compulsory education reduces migration. On the other hand, introducing foreign languages into compulsory school curricula and thus improving language skills, leads to more migration.

In our empirical analysis we exploit that education laws change over time, and hence in each country some cohorts face different lengths of compulsory schooling than other cohorts as well as distinct policies regarding compulsory foreign language classes. In particular, our empirical strategy compares migration decisions of: (i) different cohorts from the same country who were exposed to different educational polices due to policy changes, (ii) identical cohorts from different countries who were exposed to different educational policies because of differences in legislation in the two countries. In the case of foreign language classes we add an additional dimension, and we also compare different destination countries; i.e. (iii) that the same cohort in the same country was exposed to languages of some destination countries but not others. For our analysis we use Eurostat data on recent migration flows across European countries. To the best of our knowledge this is the only source that provides migration flows disaggregated by cohorts. Using mostly documentation from the European Commission's Education, Audiovisual and Culture Executive Agency, we create a novel data base on the introduction of foreign language classes into compulsory school curricula in 31 European countries in recent decades. We rely on a number of other sources such as Brunello, Fort and Weber (2009), Garrouste (2010), Hörner *et al.* (2007), and Murtin and Viarengo (2011) to complete our data base with information on changes to the length of compulsory schooling in each country in recent decades.

Controlling for economic variables (unemployment rates by cohort) in countries of origin and destination, the presence of other co-nationals, and total population by age group, we find that increasing compulsory schooling by one year reduces the number of emigrants from a country by almost 10 per cent. Introducing a foreign language into compulsory school curricula, on the other hand, almost doubles the number of emigrants to the country where the language is spoken and increase the total number of emigrants by 20 per cent. Our results are robust to a variety of alternative specifications that include historical variables (years lived under communist rule), exclude certain potentially determinant countries, or control for cohorts' years of compulsory schooling in the country of destination.

A number of recent studies use changes in education laws and related policies to instrument for education choices when estimating the causal effect of education on within- country migration4[5]. The before-mentioned paper by Machin, Salvanes and Pelkonen (2012) uses a change in compulsory schooling laws in Norway and finds education to increase internal mobility. State autonomy and geographical distance makes migration across US states more similar to our analysis of migration across European countries. Results for the effect of education on state-to-state migration in the US are mixed. Malamud and Wozniak (2010a) use the risk of being drafted for the Vietnam War as an instrument for college-level education and estimate a positive causal effect of education on migration. Results in their working paper version (Malamud and Wozniak 2010b) show that when instrumenting education by quarter of birth, the estimates turn negative but not significant. As the authors suggest, if the impact of additional educational attainment on migration differs for individuals with different baseline educational attainments, such contrasting results might arise. Similar to our analysis and findings, McHenry (2013) uses differences in changes to the minimum school leaving age across US states and shows that for low levels of education, additional educational attainment has a negative impact on state-to-state migration.

In the context of international migration, most studies use observed educational attainment, and many focus on the case of Mexican-US migration. Results from these studies range from negative self-selection of immigrants with individuals from the bottom of the skill distribution being more likely to migrate (Fernandez-Huertas Moraga 2011), to positive self-selection (Chiquiar and Hanson 2005), to a u-shaped relationship (Caponi 2010). McKenzie and Rapoport (2010) find the effect of education on Mexican-US migration to depend on the size of networks, with larger (smaller) networks attracting disproportionately more uneducated (educated) individuals.

Regarding international migration and foreign language prociency, the existing literature mainly focuses on two important aspects: its determinants and its consequences for migrants. With respect to the latter, findings by Chiswick and Miller (2010), Dustmann and Fabbri (2003), and Gonzalez (2005) show that immigrants' accomplishments in a host country's labor market depend positively and to a great extent on language skills. Bleakley and Chin (2010) find negative effects of language proficiency on fertility and marriage. Similar to the current paper, Lleras-Muney and Shertzer (2015) consider changes in education policies, among others compulsory schooling laws and the imposition of English as language of instruction. The authors find no effect of theses policies on immigrant assimilation in the US between 1910-1930. Regarding determinants of language proficiency, Chiswick [2008] points out that three aspects: (i) exposure (not being married before migration, not living in an enclave), (ii) efficiency (young age, higher education), and (iii) economic incentives (length of expected stay) positively influence the likelihood that an immigrant acquires proficiency in the host country's language. We propose a different

[5] Using data on educational attainment entails problems of reverse causality if individual decisions on education are influenced by the desire to migrate. For instance, McKenzie and Rapoport (2011) find that Mexican boys from a household with international migration experience are more likely to drop out of school.

perspective that has received relatively little attention so far: how *ex-ante* language skills influence individuals' decisions to migrate. Among the few related works are Adsera and Pytlikova [2015] who try to explain migration flows to different OECD countries using linguistic distances to measure the ease of learning a host country's language. The analysis on foreign language proficiency and migration in the current paper is similar to the one we carried out in Aparicio Fenoll and Kuehn (2014), but includes more countries, additional years, and a broader set of controls.

In contrast to most literature that uses changes in education policies as a means to investigate the effect of educational attainment on migration, the current paper directly tries to address the question: "How do changes in education policies affect migration?". While effects of education policies most likely operate through changes in individuals' educational attainment, our question is different and does not require the use of IV estimation. Our strategy allows for education policies to have general equilibrium effects that go beyond their impact on the aggregate level of education. For instance, increasing the length of compulsory schooling requires additional resources which may reduce public expenditure in other categories while improving labor market opportunities for teachers. Both aspects in turn could affect migration. In order to make sure that changes to the length of compulsory schooling affect aggregate education levels as expected, we check that these policies effectively translated into changes in educational attainment. To this end, we regress the length of compulsory schooling during the time a particular cohort was in school on measures of the cohort's average years of schooling, and as expected we find a robust positive relationship. Unfortunately, lack of cohort data on foreign language proficiency across countries prevents us from carrying out a similar test for the effect of foreign language classes during compulsory schooling on language proficiency. However, the fact that 68 per cent of Europeans obtained their foreign language skills at school (Eurobarometer 2012) provides us with confidence that also these policies were effective at achieving their means. While focusing on the impact of education policies on migration, the current paper thus also sheds light on the more general question of the effect of educational attainment on migration.

To the best of our knowledge the current paper is the first to analyze how education policies affect international migration in a multi-country setting. This is important because findings regarding the relationship between education policies and internal mobility cannot simply be extrapolated to the context of international migration. Within a country, educational attainment is easily transferable, but education obtained in one country might not be fully recognized in another country (see Chiswick 2008 or Greenwood and McDormell 1991). To the best of our knowledge we are also the first to study the impact of acquired language proficiency during compulsory education on migration. The European setting is ideal for our analysis. Basically unrestricted mobility allows us to isolate the role of education policies from migration restrictions. Outside of Europe, many countries tend to place stricter limits on the entry of low educated individuals compared to highly educated individuals, making it difficult to disentangle the effect of education policies from migration restrictions. Moreover, speaking a foreign language determines the degree to which human capital is transferable across countries, and the large variety of different languages in Europe provide a context where language proficiency is important for migration. The remainder of this paper is organized as follows: Section 2 presents the model, Section 3 describes our data. In Section 4 we present our estimation strategy. Section 5 presents and discusses our results, and Section 6 concludes.

2. Model

According to the traditional framework of the Roy model (1951) applied to the context of migration (see Borjas 1987), individuals decide to migrate upon comparing their expected incomes in origin and destination countries. Differences in returns to education across countries determine whether relatively more low or more highly educated individuals migrate. However, a simple Roy model (1951) with wages that increase monotonously in education and migration costs which are independent of an individual's education cannot generate the observed u-shaped relationship between education and migration displayed in Figure 1.2. A model that allows for heterogeneous effects of education on migration and that has the potential to generate the observed pattern is the one suggested by Stark (1991)[6]. We adapt his model and illustrate how education reforms which increase educational attainment or improve foreign language proficiency can affect migration.

Consider two countries, one rich R and one poorer country P. Expected wages in each country depend on the individuals' level of education (θ) in the following way

$$W_R(\theta) = r_0 + r_1\theta$$
$$W_P(\theta) = p_0 + p_1\theta,$$

with $r_0 > p_0$ indicating that wages are higher in the rich country[7]. Parameters r_1 and p_1 represent returns to education in the rich and poorer country respectively. Language proficiency $k \in [0, 1)$ determines how migrants' expected wages compare to those of natives with the same level of education. For high enough values of k, individuals of any education level have higher expected wages in country R than in country P,

$$kW_R(\theta) > W_P(\theta).$$

However, an education obtained in one country is not automatically recognized elsewhere. In order to have an educational degree officially recognized in a foreign country, individuals have to incur in cost C that includes official translations and administrative paperwork requested by government agencies, associations, or guilds. Similar costs might also arise due to license requirements for certain professions. Such requirements do not

[6] Caponi (2010) proposes a model where transferability of human capital is limited across countries but parents migrate for a better education of their children. His model also generates a u-shaped relationship between education and migration.

[7] Note that these parameters can also incorporate aspects that affect differences in the probability of finding a job, e.g differences in unemployment rates.

only differ across countries but also across US states (see e.g. Federman, Harrington and Krynski 2006). According to Kleiner and Krueger (2010) occupational licenses are more prevalent among workers with high school education (22 per cent) and college degrees (44 per cent) than high school dropouts (12 per cent). Net income in the foreign country is hence given by $kW_R(\theta) - C$. If individuals migrate without a recognized degree they are able to earn a minimum wage that does not depend on one's education, kW_R[8].

Given certain parameter values, a u-shaped relationship between educational attainment and migration arises. The upper graph of Figure 2.3 displays the situation in which returns to education are higher in the destination country and the minimum wage is higher than the wage of low educated individuals in the country of origin. However, the minimum wage lies below the wage that medium educated individuals can earn in the country of origin. This gives rise to a u-shaped pattern. Individuals with low educational attainment – with $\theta < \theta_1$ – migrate without having invested in degree recognition and they earn minimum wage kW_R. Individuals with a medium level of education – between θ_1 and θ_2 – do not migrate and they earn $W_p(\theta)$. Finally, those with higher educational attainments ($\theta > \theta_2$) pay the costs to have their degree recognized and migrate. They earn $kW_R(\theta)–C$.

An increase in the length of compulsory schooling shifts a mass of individuals from low educational attainment towards medium educational attainment. This leads to a new distribution of education with a higher mean and smaller variance as displayed in the lower graph of Figure 2.3. Given our parameter values, this implies an increase in the share of individuals who do not migrate. The difference between the light and dark gray area indicates the additional mass of individuals who decide to stay as a consequence of the increase in the length of compulsory schooling. Everything else equal, our model predicts that an increase in the length of compulsory schooling reduces the number of individuals who migrate.

As language proficiency increases from k to k' – for instance as a consequence of the introduction of foreign languages into compulsory school curricula – migrants are able to obtain wages that are more similar to those of natives. Figure 2.4 displays what happens as language proficiency increases[9]. Expected wages are higher, both for those who migrate without degree recognition as well as for those who invest in degree recognition. However, in line with findings in literature the functional form for the relationship between wages and education is such that the increase in expected wages and hence the gains from language proficiency are larger for highly educated individuals (see McManus, Gould and Welch 1983; McManus 1985; Mora and Davila 1998; Carliner 1996).

[8] This simple static model abstains from improvements in language proficiency once the individual has migrated, nor does it allow for return-migration. Reinhold and Thom (2013) propose a model of return migration where individuals' migration decisions also take into account the potential increase in income back home, something that could in part be due to language acquisition.

[9] Without loss of generality, parameter values for Figures 2.3 and 2.4 are chosen such that $C = kr_0$.

The minimum wage increases to $k'\bar{W}_R$ and the intercept and slope of the expected wage function for recognized degrees ($k'W_R(\theta)-C$) increase as well. As a result fewer individuals decide to stay. An increasing number of high and low educated migrate. The model thus predicts that an increase in the length of compulsory schooling reduces migration, while the introduction of foreign languages into compulsory school curricula leads to more migration.

In order to fully test the model, we would need to check our assumptions about wage profiles in the destination and origin countries. However, differences in wage profiles while typically included in estimations of internal migration decisions – see Kennan and Walker (2011) for the United States – are usually not available for analyses of international migration. One exception is Bertoli, Fernandez-Huertas Moraga and Ortega (2013) who consider migration from one origin country, Ecuador, to two destination countries, Spain and the United States. Our analysis extents to migration across 31 European countries for which comparable estimates for migrants' wages profiles are not available. We can thus only test the model's predictions for migration decisions under different educational policies regarding length of compulsory schooling and foreign languages in compulsory school curricula.

FIGURE 2.3 • EFFECT OF INCREASE IN LENGTH OF COMPULSORY EDUCATION ON MIGRATION

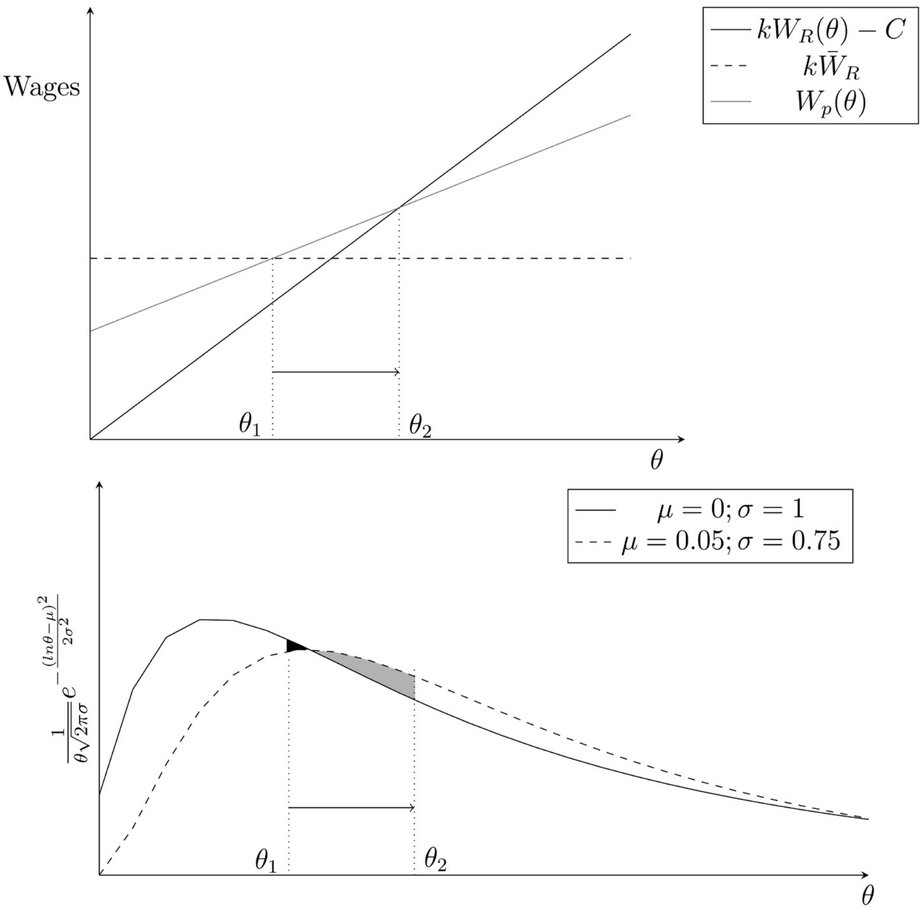
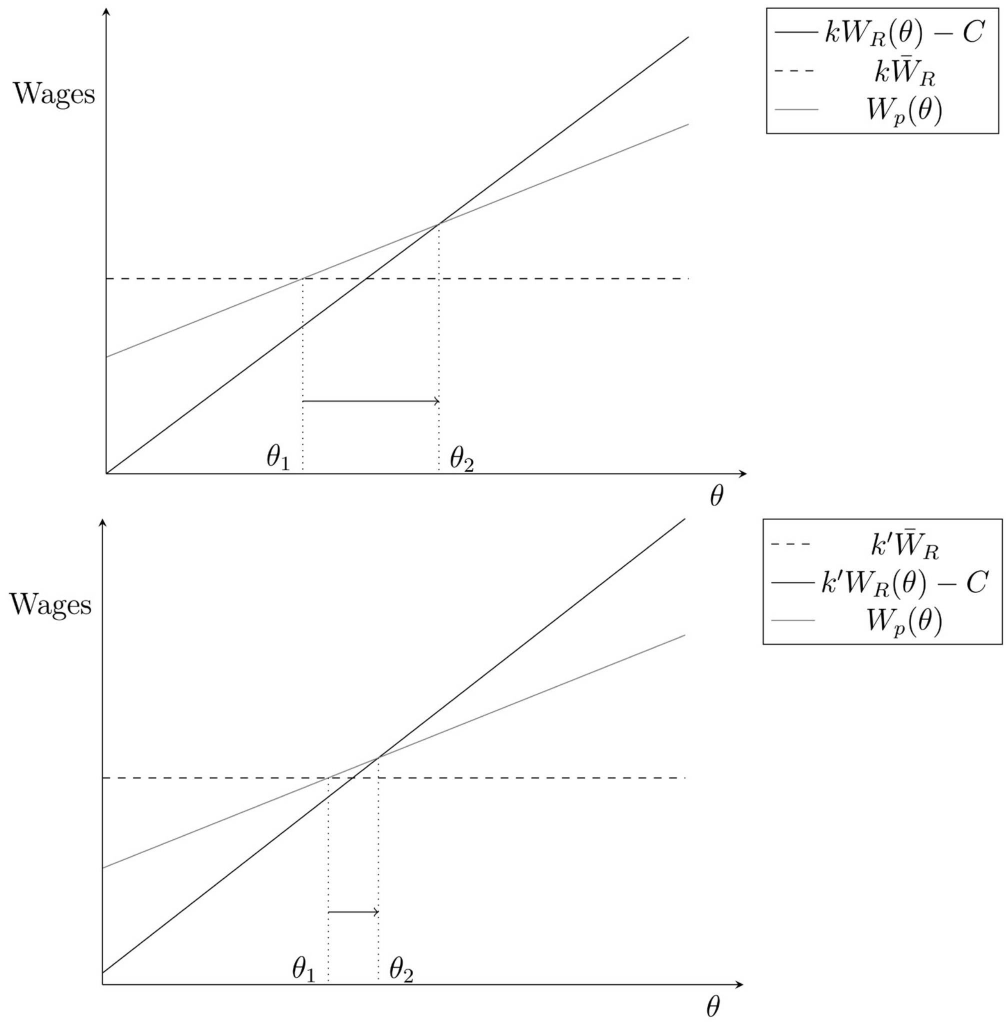


FIGURE 2.4 • EFFECT OF IMPROVEMENT IN LANGUAGE PROFICIENCY FROM K TO K' ON MIGRATION



3. DATA

For our analysis we use data from Eurostat on migration across European countries. In particular, we consider the flow in t and stock of immigrants in $t-1$ by 5-year age groups for all combinations of origin and destination countries in 2008-2012. The following 26 destination countries provide this data: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Germany, Hungary, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Netherlands, Norway, Poland, Romania, Slovakia, Slovenia, Spain, and Sweden. For Germany and Austria, missing data for 2009-2012 and 2010 respectively is complemented with data from the Statistische Bundesamt and Statistik Austria. Data for the UK come from the International Passenger Survey of the Office for National Statistics (ONS). We thus have information on 27 destination countries and 31 countries of origin – all destination countries plus France, Greece, Malta, Latvia, and Portugal. We also rely on Eurostat data for national unemployment rates by 5-year age groups.

For our analysis we consider young individuals between 25 and 44 who are most likely to migrate for work-related reasons. To avoid picking up short-term temporary migration related to studying abroad, we restrict our sample to individuals age 25 and older. Furthermore, in many countries for individuals older than 44 (born before 1964) it is unclear that language learning (or even compulsory schooling) was enforced. We construct a database with information on the required years of compulsory schooling for each cohort in each country based on four main sources: Brunello, Fort and Weber (2009), Garrouste (2010), Hörner *et al.* (2007), Murtin and Viarengo (2011). Educational reforms that changed the length of compulsory schooling during the 20th century for different cohorts generate within- and across-country variation. Table A.1 of the Appendix displays these changes and variations for our cohorts.

We create a novel database on foreign language classes in compulsory education using mainly information from the European Commission's Education, Audiovisual and Culture Executive Agency (EACEA) and the European's Commission's Directorate-General for Education and Culture. For each cohort and country, this database includes information on the starting age for studying foreign languages during compulsory education as well as on the type of languages studied. Educational reforms that have occurred during the last decades imply that some cohorts have been exposed to foreign languages during compulsory education while others have not. There are also differences in the type of foreign languages included in school curricula. In most former communist countries of Central and Eastern Europe after 1990, Russian was replaced by English as the first 12 foreign language. Nowadays, the vast majority of students in European countries studies English as their first foreign language. In many countries, studying a second foreign language is compulsory during lower secondary education. Traditionally, only German and French were offered as second foreign languages, but recently individuals in most countries can also choose Spanish and in fewer countries Italian. At present, German is more common in Central and Eastern Europe, while French tends to be taught in Southern European countries. Spanish is the third or fourth most widely taught second foreign language. These differences and changes over time generate variations within- and across-countries of origin and destination in the exposure to foreign languages. For countries, where students can choose among various foreign languages we consider all options to avoid picking up individual choices which can be endogenous to migration decisions. We also take into account that there are countries where studying a second foreign language is not part of compulsory education, and that students in Finland learn Swedish as a foreign language. Our data set contains this information by cohort and country of origin. We summarize this information in Table A.2 of the Appendix.

Table 3.1 provides summary statistics – mean, standard deviation and minimum and maximum values – for our variables. We have observations for 11,205 cells defined by the combination of origin, destination, age, and year[10]. On average, 221 individuals in each age group

[10] In total we should have 14,500 observations. Unfortunately, we were not able to complement the following missing data: destination countries France, Greece, Malta, and Latvia, data for Belgium 2008-2009, Bulgaria 2009-2011, Croatia 2009-2010, Cyprus 2009-2012, Macedonia 2009-2010, Poland 2009-2012, Portugal 2009-2012, Slovakia 2012. The remaining missing data refer to single observations, for instance for half of all destination countries migration from Liechtenstein is missing.

from each country of origin migrate each year to one of the destination countries. However, we observe a lot of variation in these migration flows. There was no migration from Cyprus to Estonia in 2010, while in 2008, 29,250 individuals age 25-29 migrated from Romania to Italy. Average years of compulsory schooling are 9.1, ranging from 6 (for older cohorts in most countries) to 13 for younger cohorts in Germany. Around 11.3 per cent of our observations – cells defined by the combination of origin, destination, age, and year – were exposed to compulsory foreign language classes in the language of destination countries. Observations are distributed homogeneously across age groups. We have slightly more observations for 2008 than for 2009-2012. Regarding differences in unemployment rates, measured one year before migration, we observe a maximum difference of 33 percentage points between unemployment rates in Norway and Greece for individuals age 25-29 in 2012. Also measured one year before, the average number of immigrants of a certain age group from a certain country of origin is around 1,797, i.e. more than eight times the average annual inflow.

Table 3.1 • Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
Imm flow origin-destination by age	221.199	1,015.603	0	29,250
Years of compulsory education	9.096	1.184	6	13
Exposed to foreign language	0.113	0.311	0	1
Age group 25-29	0.252	0.434	0	1
Age group 20-34	0.252	0.434	0	1
Age group 35-39	0.249	0.432	0	1
Year: 2008	0.224	0.417	0	1
Year: 2009	0.183	0.387	0	1
Year: 2010	0.193	0.395	0	1
Year: 2011	0.195	0.396	0	1
Diff unemp origin-destination by age	0.474	5.949	-27.2	33.1
Stock imm origin-destination by age	1,796.978	10,279.171	0	324,571
Stock population in origin by age	1,182,331.963	1,603,305.056	2,224	7,176,550
Years under communist rule	4.845	7.474	0	23.5

N = 11,205; Differences in unemployment rates and the stock of immigrants refer to years $t-1$, i.e 2007, 2008, 2009, 2010, 2011. Sources: Eurostat, Statistisches Bundesamt, Statistik Austria, ONS, UN Data, Eurybase, European Commission's Education, Audiovisual and Culture Executive Agency (EACEA), European's Commission's Directorate-General for Education and Culture, Brunello, Fort and Weber (2009), Garrouste (2010), Hörner *et al.* (2007), Murtin and Viarengo (2011), etc.; own calculations.

However, while in 2010 there were no individuals from Bulgaria residing in Liechtenstein, there were 324,571 immigrants from Poland aged 40-44 living in Germany in 2011. On average, there are a little over 1 million inhabitants per age group in each country of origin, ranging from only 2,224 individuals of age 25-29 in Luxembourg and Liechtenstein, to more than 7 million British and Germans of age 40-44. Finally, we calculate the average number of years lived under communist rule by a cohort as the difference between 1990 and the cohort's birth year.

4. Estimation Strategy

Our identification strategy makes use of all three dimensions of variation in the data, comparing individuals across age, countries of origin, and time. We could have estimated the effects of education policies on migration by only comparing individuals across age, considering different cohorts from the same country who were affected by different compulsory schooling laws. However, such an estimation is affected by differences in the propensities to migrate by age. Another alternative would have been to compare individuals of the same cohort from different countries of origin. However, nationals of different countries have different propensities to migrate, independently of education policies. A third approach would have consisted in observing individuals of a certain age and country of origin at different points in time, using the fact that they were affected by different compulsory schooling laws. However, we only have data on migration flows for five years, and even if we disposed of additional data, migration patterns change over time. We improve upon these approaches by combining them all. Using fixed effects, this strategy allows us to control for confounding factors that vary with age, time, and country of origin, and their pairwise combinations (age and time, age and country, time and country). Moreover, in our estimation of the impact of compulsory foreign language classes we introduce a fourth dimension, using the fact that our explanatory variable varies by destination country. We hence compare the propensity to migrate to different destination countries, and we attribute the difference in migration flows between the destination country where the taught language is spoken and other countries to the impact of foreign language classes. As a result, our estimated coefficients result from refined comparisons of cohorts, and they are robust to the potential influence of a long list of unobserved factors.

To assess the impact of education policies on migration we estimate two separate models. Regarding our first model, we estimate the effect of years of compulsory education on the number of migrants in a cohort. We assume the following linear form for the relationship between the two variables:

$$\begin{aligned} M_{a,o,d,t} &= \alpha_0 + \alpha_1 CS_{a,o,t} + \alpha_2 D_a + \alpha_3 D_o + \alpha_4 D_d + \alpha_5 D_t + \alpha_6 D_{a,o} + \alpha_7 D_{a,d} + \\ &\quad + \alpha_8 D_{a,t} + \alpha_9 D_{o,d} + \alpha_{10} D_{o,t} + \alpha_{11} D_{d,t} + \alpha_{12} X_{t-1} + \epsilon_{a,o,d,t} \end{aligned} \tag{4.1}$$

where M is the number of immigrants of age a from country o going to country d in year t. CS denotes the number of years of compulsory schooling faced by individuals of age a in year t, and D_s with $s = a, o, d, t$ are dummies for age, country of origin, country of destination, and year. Our basic model includes all four dummy variables and two interaction terms for age and year and country of origin and destination.

We then expand the model and include all simple interactions as well as certain double interactions of dummy variables. For instance, we add the interaction term $D_{d,o,t}$ between country of origin, country of destination, and year. This term accounts for pull and push factors between country pairs that change over time and that affect individuals regardless of their age. Including these dummy variables is equivalent to including control variables from typical gravity models like differences in the share of young individuals in the labor force, female labor force participation rates, or average wage differentials (see e.g. Ortega and Peri 2009 or Lewer and Van den Berg 2008). In our context, including these terms in the estimation is important for two reasons: (i) In 2009 and 2011, work restrictions in some countries for nationals of Central European countries that joined the EU in 2004 and 2007 respectively, were finally lifted. (ii) Four countries in our sample (Croatia, Liechtenstein, Macedonia, Norway) did not belong to the EU during 2008-2012. Norway and Liechtenstein belong to the Schengen area which guarantees free mobility since 2001 and 2011 respectively. Croatia joined the EU in 2014, and Macedonia is an EU candidate country, and since 2009 its residents can travel visa-free to the Schengen area. Note that when in place, these restrictions applied to individuals of all ages. We also include the interaction term $D_{d,a,t}$ between destination country, age group, and year, to take into account any age-specific changes in the labor demand of the destination country. Moreover, to control for network effects and economic factors we include as lagged control variables by age group (X_{t-1}) total population in country of origin, stock of immigrants settled in the destination country, and differences in the unemployment rates between the destination country and the country of origin.

For countries where the length of compulsory schooling did not change during the time our cohorts were in school, the variable $CS_{a,o,t}$ is a constant. As a result, the corresponding dummy variable $D_{a,o}$ will not be identified. However, as long as identifying this dummy variable is not the focus of our analysis, this will not pose a problem for our estimation. Following Bertrand, Duflo and Mullainathan (2004), we cluster standard errors at the destination-origin-age level to allow for serial correlation in migration flows over time.

Our second model estimates the number of migrants as a function of exposure to compulsory foreign language classes in the language of the destination country[11]. We assume the following linear form for the relationship between the two variables:

$$M_{a,o,d,t} = \beta_0 + \beta_1 L_{a,o,d,t} + \beta_2 D_a + \beta_3 D_o + \beta_4 D_d + \beta_5 D_t + \beta_6 D_{a,o} + \beta_7 D_{a,d} + \\ + \beta_8 D_{a,t} + \beta_9 D_{o,d} + \beta_{10} D_{o,t} + \beta_{11} D_{d,t} + \beta_{12} X_{t-1} + \epsilon_{a,o,d,t} \quad (4.2)$$

where L is a dummy variable that denotes exposure to compulsory language classes in at least one of the official languages of country *d*. All other variables are as defined before.

Only some foreign languages are studied during compulsory education in European schools, and hence the set of foreign languages considered includes English, German, French, Spanish, and Italian[12]. For destination countries where neither English, German, French, Spanish, or Italian are ocial languages, we set $L_{a,o,d,t} = 0$ for all *t, a, o.* As a result, in our model specification that in-

[11] Limited reliability of data on years of exposure to foreign language classes is the main reason why we do not consider such a refinement of our dependent variable. Furthermore, comparability across countries would require adjustments for hours taught per week as well for linguistic distance between language of destination and origin countries.

[12] Even though Russian is the most widely taught second foreign language in Latvia, Estonia, and Lithuania, we ignore this option given that we do not have data on migration flows to Russia.

cludes triple interactions, dummy variables $D_{d,o,t}$ or $D_{d,a,t}$ are not going to be identied. As mentioned before, as long as identifying these dummy variables is not our main interest, this will not pose a problem. Again, we cluster standard errors at the destination-origin-age level to allow for serial correlation in migration flows over time.

4.1 Endogeneity Concern

There might exist some concern that education reforms could be endogenous to migration. Endogeneity could arise for two reasons: (i) reverse causality: if somehow cohort-specific migration patterns in 2008-2012 determined education reforms implemented in the past when those cohorts were in school. (ii) Omitted variables: if determinants of cohort-specific migration patterns (e.g differences in cohort-specic labor market conditions between origin and destination countries) persisted over time, and if they influenced reforms that were implemented when our cohorts were in school.

Regarding the first concern: (i) Education reforms are predetermined with respect to migration patterns in 2008-2012. Still, migration patterns could be highly persistent over time. However, education reforms could at most be driven by aggregate migration flows. It is highly unlikely that they are determined by differences in the number of migrants by cohort. Hence, origin-year fixed effects and origin-destination-year fixed effects pick up any effect of migration persistence in our estimations on the effect of number of years of compulsory schooling and foreign language classes, respectively. (ii) From a political economy point of view, migration flows are unlikely to influence education policies. Governments design their education policies focusing on the median voter who stays, instead of targeting those who migrate. Moreover, the time that passes from the moment education reforms are implemented to the time that students finish their compulsory education and enter the labor market – be it at home or abroad – is likely to exceed governments' mandates. Hence, as governments might not be able to reap the potential fruits in terms of more or less migration, migration flows or brain drain concerns are very unlikely determinants of policies affecting compulsory education[13].

Regarding the second concern: To proxy labor market conditions, in our estimations we control for differences in cohort-specific unemployment rates in the year before migration, and our estimated coefficients remain unchanged. This suggests that differences in labor market conditions between origin and destination countries are not driving education reforms implemented in the past. One could think that unemployment rates at the time of the reforms could be a relevant omitted variable, however those are unlikely to affect migration patterns in 2008-2012, in particular once controlling for contemporaneous unemployment. In general, in order to address both concerns, one would like to know

[13] The only example known to us of a government explicitly providing training such that its citizens become better migrant workers is the training of nurses in the Philippines, see Lorenzo *et al.* (2007). The effect of such specialized training of adult workers on migration is much more immediate than the one resulting from education reforms regarding compulsory schooling.

more about the determinants of education reforms. To the best of our knowledge there does not exist any established theory on the political economy of education reforms, but increasing the length of compulsory education or introducing foreign language classes into compulsory school curricula requires additional resources: teachers, facilities, etc. The actual implementation of those reforms hence depends on the availability of resources. In our regressions, fixed effects pick up business cycle phases which determine the availability of resources to a large extent (e.g. tax revenues, social expenditures).

Furthermore, there is a long tradition in the literature of using compulsory schooling laws as exogenous shifters of educational attainment. Since the seminal paper by Angrist and Krueger (1991), such laws have been used in studies of the causal relationship between education and many different outcomes, such as earnings (Harmon and Walker 1995), health (Brunello, Fabbri and Fort 2013), and citizenship (Milligan, Moretti and Oreopoulos 2004). More relevant to us and as discussed before, changes in the length of compulsory education have also been used in analyses of the impact of education on internal migration in Norway (Machin, Salvanes and Pelkonen 2012) and the United States (Malamud and Wozniak 2010b; McHenry 2013).

5. Results

We first test whether changes to the length of compulsory education have any impact on the actual number of years of schooling. Only if those changes are actually enforced and individuals are not already staying in school beyond the minimum years required by law, can we expect an effect of compulsory schooling on indirectly related outcomes like migration. We hence regress average years of schooling as measured by Barro and Lee (2010) for different age groups for 2010 on our measure of years of compulsory schooling as determined by each country's education policy. Unfortunately, we cannot exploit changes over time because Barro and Lee (2010) only provide data every five years[14]. We also include dummy variables for age group, country of origin, and destination and their simple interactions in this regression. Table A.3 of the Appendix shows the results from this regression. The estimated coefficients indicate that policies that increased the length of compulsory schooling were effective in increasing average years of education for the affected cohorts.

We then turn to our empirical analysis regarding the impact of additional years of compulsory schooling on the propensity to migrate. Results from our first model as defined in Equation 4.1 are displayed in Table 5.1. The first column corresponds to the basic regression that includes dummy variables for year, age group, and countries of origin and destination, two interaction terms for age and year and country of origin and destination, as well as our lagged control variables for unemployment, stock of immigrants and population by age group. In column 2 we add all simple interactions. In column 3 we include a triple interaction (destination by origin by year). Column 4 presents results for the most complete specification which also includes the triple interaction of destination, age group,

[14] Barro and Lee (2010) provide data for all countries included in the main estimations, except Liechtenstein.

and year. Our coefficient of interest is negative, significant, and very stable across specifications. An additional year of compulsory schooling decreases the number of immigrants from the affected cohort who migrate in a given year to one specific destination country by 21 individuals. This implies a reduction of 9.7 per cent with respect to the overall mean, and 14.6 per cent with respect to a mean considering only countries providing identification, i.e. those that carried out education reforms that affected the cohorts in our sample.

Table 5.1 • Migration and years of compulsory schooling

Years of compulsory schooling	(1)	(2)	(3)	(4)
	-24.618	-31.383	-32.706	-31.295
	(15.301)	*(15.238)***	*(14.346)***	*(14.497)***
Obs.	11,205	11,205	11,205	11,205
R^2	0.683	0.713	0.928	0.93

The dependent variable is the number of immigrants. The variable years of compulsory schooling refers to the average number of years of compulsory schooling faced by the corresponding cohort. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year-fixed effects, age indicators, binary variables for each pair of origin and destination countries, dummies for each combination of age and year, a variable for differences in lagged age-specific unemployment rate between origin and destination countries, the stock of co-nationals from each age group in the destination country in the previous period, and the size of the age group in the origin country. Errors are clustered by origin-destination-age.

Table 5.2 – similarly structured as Table 5.1 – contains the estimation results of our model that considers the effect of compulsory foreign language classes on migration. In particular, we consider how having been exposed to English, French, and German during compulsory education raises the odds of migrating to the UK, Ireland, Belgium, Germany, or Austria as compared to the odds of migrating to any other European country. For younger individuals in countries like Bulgaria, Finland, France, and Netherlands we also consider if having been exposed to Spanish increases the odds of migrating to Spain. Finally, for individuals in countries like Malta, Slovenia, and Austria we also consider if having been exposed to Italian increases the odds of migrating to Italy. Our results show that this is the case. The coefficient of interest remains stable even after controlling for simple interactions of country of destination, country of origin, age group, and year effects and some second order interactions. We find that exposure to foreign language classes during compulsory education almost doubles migration to the country where the language is spoken. The number of individuals of a cohort that migrate to this country increases by 392 individuals per year, 177 per cent higher than the mean. Our estimated coefficient could in principle be driven by two aspects: (i) increased migration and (ii) redirected migration towards countries where taught languages are spoken ("substitution effect")[15]. In order to

[15] A model as the one in Bertoli, Fernandez-Huertas Moraga and Ortega (2013) where individuals first

compare this estimate to the previous one for compulsory education, one may wonder how much it represents in terms of overall migration. Given that we cannot disentangle the two driving forces, we can only calculate an upper bound for the overall effect that corresponds to a situation when substitution effects are absent. Considering that only 11.3 per cent of our observations were exposed to compulsory foreign language classes in the language of destination countries, the total number of emigrants increases by 20 per cent.

Table 5.2 • Migration and compulsory foreign language classes

Foreign Language Classes	(1)	(2)	(3)	(4)
	575.208	459.246	425.323	421.887
	*(152.695)****	*(166.262)****	*(166.684)***	*(170.595)***
Obs.	11,205	11,205	11,205	11,205
R^2	0.688	0.727	0.93	0.932

The dependent variable is the number of immigrants, the variable foreign language classes identifies the cohorts from the country of origin who were exposed to learning the language of the country of destination during compulsory schooling. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year-fixed effects, age indicators, binary variables for each pair of origin and destination countries, dummies for each combination of age and year, a variable for differences in lagged age-specific unemployment rate between origin and destination countries, the stock of co-nationals from each cohort in the destination country in the previous period, and the size of the age group. Errors are clustered by origin-destination-age.

5.1 Robustness Checks

If different countries simultaneously increased the length of compulsory schooling, less migration could be driven by more education in the origin country as well as more education in the destination country. The latter might result from increased competition and lower wages for high-skilled jobs. Hence, we also test for such effects by including years of compulsory schooling in the destination country into our regressions. The coefficient for this variable is only significant in our first most simple specification, and the coefficient for "years of compulsory schooling" in the origin country remains unaltered; see Table A.4 of the Appendix.

As mentioned before, young individuals in most European countries study English as their first foreign language during compulsory education. Hence, migration to the UK could be of particular importance for our estimations. Data for the UK, different from our other data (Eurostat, Statistisches Bundesamt, Statistik Austria) are not based on registers but are estimated based on international passenger flows. Hence, there might be some concern regarding the fact that results could be exclusively driven by these data. However, when run-

decide to migrate and then specify their destination country could produce such substitution effects.

ning our regressions excluding observations for the destination country UK, the coefficient for "foreign language classes" remains unaltered; see Table A.5 of the Appendix.

In Central and Eastern Europe – with the exception of Croatia, Macedonia, Romania, and Slovenia where Russian was not the first compulsory foreign language – the change from Russian to English as the first foreign language was driven by the end of communism, which in itself had important implications for migration flows. Given that we consider migration in years 2008-2012, most of the initial emigration boom is likely to have ebbed out. Even if that were not the case, in our estimations we compare migration decisions of individuals who were and those who were not exposed to English as a foreign language.

If the end of communism was still the main driving force for migration in 2008-2012, then – controlling for differences in age – we should not observe any differences in migration decisions between the two groups. For instance, among two cohorts from the same ex-communist country, one born in 1980 and another one born in 1975, the former was exposed to English as a foreign language while the latter was not. If individuals from both cohorts migrated to the UK or Ireland, origin-destination-fixed effects would capture their decision, and it would not add to our estimated effect of language proficiency on migration. Only in case individuals from the younger cohort, but not from the older cohort migrated to the UK or Ireland would we attribute their migration decisions to the newly acquired English skills. In order to address any remaining concerns, we include a control variable for the number of years a cohort lived under communist rule into our estimations. Our results remain unchanged – see Tables A.6 and A.7 of the Appendix.

We also carry out a formal check addressing the exogeneity of educational reforms. For the case of a nation-wide education reform that was implemented sequentially by Norwegian municipalities, both Machin, Salvanes and Pelkonen (2012) and Black, Devereux and Salvanes (2005) provide a test for the exogeneity of this timing. They suggest running a regression of the birth year of affected cohorts on a variety of socio-economic variables (income, labor force participation, educational attainment, election outcomes, etc.) measured around the time of the reform. The authors of both papers conclude that county fixed effects turn out to be the only significant variables in these regressions. As those same fixed effects are included in their main regressions, this dependence does not pose any problem. We run a similar regression of a dummy variable for reform that varies at the year and country of origin level on a variety of potentially related variables, as well as year and country dummies. As potentially related variables we consider available data on demographics (population growth), urban development (percentage rural population), education (average years of schooling), economic development (value added share of manufacturing, and agriculture) and business cycle (unemployment rate, GDP per capita). Tables A.8 and A.9 of the Appendix show the results from this estimation for reforms regarding changes in length of compulsory schooling and foreign language classes respectively. Governments' educational reforms regarding foreign language classes in compulsory education or the length of compulsory schooling do not seem to relate systematically to changes in population growth, urban or economic development, education, or the business cycle. As in the two above-mentioned papers, only few fixed effects are signicant. Moreover, the proportion of the variance attributable to the regressors is low.

6. Conclusion

Previous literature has used education reforms to test whether more education is associated with more or less within-country mobility. Results have been mixed. We consider an international context with basically unrestricted migration – Europe – and test for the direct effect of education policies on migration. We show that increases in the length of compulsory education reduce the propensity to migrate across European countries. One additional year of compulsory education reduces migration by almost 10 per cent. We also show that the introduction of foreign languages into compulsory school curricula on the other hand, increases migration. In particular, we find that acquiring foreign language proficiency during compulsory education almost doubles the number of individuals who migrate to the country where the language is spoken.

One of the top priorities of the European Union (EU)'s 2020 agenda is to improve educational outcomes. Education policies that lead to a more educated and better prepared workforce are essential for future growth and job creation. At the same time, labor mobility is one of the main EU objectives, and foreign language proficiency, key for human capital transferability across countries is ranked a chief concern in the Barcelona objective of 2002. Our results suggest that education policies aimed at increasing educational attainment and foreign language proficiency may have opposite effects on migration. Increasing educational attainment while reducing differentials in national unemployment rates across Europe thus requires coordinated education and labor market policies[16]. For lower levels of education, our results show that governments can be fairly confident that more years of compulsory schooling are unlikely to be lost to brain drain. Our results do not extent to higher levels of education, but within the context of our theoretical model we can conjecture that initiatives like the Bologna process that makes university degrees across Europe comparable can help to reduce degree recognition costs, increase returns to migration and foster mobility. Recently, the EU Commission has proclaimed the ambitious goal of enabling all EU citizens to communicate in two languages other than their mother tongue. If the EU wants to seriously promote a unified labor market, it should strengthen teaching of foreign languages in compulsory education, and in particular of those languages spoken in countries with strong labor markets (German instead of Spanish).

7. References

Adsera A. and Pytlikova M. (2015), *The role of language in shaping international migration*, «Economic Journal», 125, 586, pp. F49-F81

Angrist J.D. and Krueger A.B. (1991), *Does compulsory school attendance affect schooling and earnings?*, «Quarterly Journal of Economics», 106, 4, pp. 979-1014

[16] Boldrin and Canova (2001) argue that EU policies aimed at achieving convergence in economic conditions across Europe seem to discourage migration at the same time.

Antolin P. and Bover O. (1997), *Regional migration in Spain: The effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections*, «Oxford Bulletin of Economics and Statistics», 59, pp. 215-235

Aparicio Fenoll A. and Kuehn Z. (2014), *Does foreign language proficiency foster migration of young individuals within the European Union?*, IZA, Discussion Paper n. 8250

Archimedes Foundation (2010), *Higher education in Estonia*, Estonian Academic Recognition Information Centre, 4th edition, Tallin

Barro R. and Lee J.-W. (2010), *A new data set of educational attainment in the world, 1950-2010*, «Journal of Development Economics», 104, pp. 184-198

Bartz K. and Fuchs-Schündeln N. (2012): *The role of borders, languages, and currencies as obstacles to labor market integration*, «European Economic Review», 56, 6, pp. 1148-1163

Belot M. (2007), *Why is employment protection stricter in Europe than in the United States?*, «Economica», 74, pp. 397-423

Bertoli S., Fernández-Huertas Moraga J. and Ortega F. (2013), *Crossing the border: Self-selection, earnings and individual migration decisions*, «Journal of Development Economics», 101, pp. 75-91

Bertrand M., Dufloo E. and Mullainathan S. (2004), *How much should we trust differences-in-differences estimates?*, «Quarterly Journal of Economics», 119, 1, pp. 249-275

Black S.E., Devereux P.J. and Salvanes K.J. (2005), *Why the apple doesn't fall far: Understanding intergenerational transmission of human capital*, «American Economic Review, 95, 1, pp. 437-449

Bleakley H. and Chin A. (2010), *Age at arrival, English proficiency, and social assimilation among US immigrants*, «American Economic Journal: Applied Economics», 2, 1, pp. 165-192

Boldrin M., and Canova F. (2001), *Inequality and convergence in Europe's regions: reconsidering European regional policies*, «Economic policy», 16, 32, pp. 206-253

Borjas G.T. (1987), *Self-selection and the earnings of immigrants*, «American Economic Review», 77, 4, pp. 531-553

Braham R.L. (1972), *Education in Romania: A decade of change*, Office of Education, U.S. Department of Health, Education and Welfare, Washington D.C.

Brücker H., Capuano S. and Marfouk A. (2013), *Education, gender and international migration: insights from a panel-dataset 1980-2010*, mimeo, German Institute for Employment Research (IAB)

Brunello G., Fort M. and Weber G. (2009), *Changes in compulsory schooling, education and the distribution of wages in Europe*, «Economic Journal», 119, 536, pp. 516-539

Brunello G., Fabbri D. and Fort M. (2013), *The causal effect of education on body mass: Evidence from Europe*, «Journal of Labor Economics», 31, 1, pp. 195-223

Caponi V. (2010), *Heterogeneous human capital and migration: Who migrates from Mexico to the US?*, «Annals of Economics and Statistics», 97-98, pp. 207-234

Carliner G. (1996), *The Wages and Language Skills of U.S. Immigrants*, NBER wp n. 5763

Chiquiar D. and Hanson G.H. (2005), *International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States*, «Journal of Political Economy», 113, 2, pp. 239-281

Chiswick B.R. (2008), *The Economics of Language: An Introduction and Overview*, IZA, Discussion Paper n. 3568.

Chiswick B.R. and Miller P.W. (2010), *Occupational language requirements and the value of English in the US labor market*, «Journal of Population Economics», 23, 1, pp. 353-372

Directorate General for Education and Culture (2001), *Foreign Language Teaching in Schools in Europe*, Eurydice The Information Network on Education in Europe

Dustmann C. and Fabbri F. (2003), *Language proficiency and labour market performance of immigrants in the UK*, «Economic Journal», 113, 489, pp. 695-717

Education, Audiovisual and Culture Executive Agency (2010), *Organisation of the education system in Liechtenstein 2009/10*

– (2012), *Key data on teaching languages at school in Europe 2012*

Eurobarometer (2012), *Europeans and their Languages*, Special Eurobarometer 386, Directorate General for Education and Culture, European Commission

Eurydice (1997), *Supplement to the study on the structures of the education and initial training systems in the European Union: The situation in Bulgaria, the Czech Republic, Hungary, Poland, Romania and Slovakia*, Eurydice, The Information Network on Education in Europe

Federman M.N., Harrington D.E. and Krynski K.J. (2006), *The impact of state licensing regulations on low-skilled immigrants: The case of Vietnamese manicurists*, «American Economic Review», 96, 2, pp. 237-241

Fernández-Huertas Moraga J. (2011), *New Evidence on Emigrant Selection*, «Review of Economics and Statistics», 93, 1, pp. 72-96

Gálvez C.M., Arrimadas Gömez I. *et al.*, *La enseñanza de lenguas extranjeras en España*, Ministerio de Educacion Cultura y Deporte

Garrouste C. (2010), *100 years of educational reforms in Europe: A contextual database*, JRC Scientific and Technical Reports, European Commission

Gáková Z. and Dijkstra L. (2008), *Labour mobility between the regions of the EU-27 and a comparison with the USA*, Regional Focus, European Union, Regional Policy n. 2.

González L. (2005), *Nonparametric bounds on the returns to language skill*, «Journal of Applied Econometrics», 20, 6, pp. 771-795

Gray R.D. and Atkinson Q.D. (2003), *Language-tree divergence times support the Anatolian theory of Indo-European origin*, «Nature», 426, pp. 435-439

Greenwood M.J. and McDowell J.M. (1991), *Differential Economic Opportunity, Transferability of Skills, and Immigration to the United States and Canada*, «Review of Economics and Statistics», 73, 4, pp. 612-623

Harding R.M. and Solkal R.R. (1988), *Classification of the European language families by genetic distance*, «Proceedings of the National Academy of Science USA», 85, 23, pp. 9370-9372

Harmon C. and Walker I. (1995), *Estimates of the economic return to schooling for the United Kingdom*, «American Economic Review», 85, 5, pp. 1278-1286

Hörner W., Döbert H. *et al.* (eds) (2007), *The Education Systems of Europe*, Dordrecht, Springer

Jauer J., Liebig T. *et al.* (2014), *Migration as an adjustment mechanism in the crisis? A comparison of Europe and the United States*, OECD Social, Employment and Migration Working Paper n. 155

Kennan J. and Walker J.R. (2011), *The effect of expected income on individual migration decisions*, «Econometrica», 79, 1, pp. 211-251

Kleiner M.M. and Krueger A.B. (2010), *The prevalence and effects of occupational licensing*, «British Journal of Industrial Relations», 48, 4, pp. 676-687

Lewer J. and Van den Berg H. (2008), *A gravity model of immigration*, «Economics Letters», 99, 1, pp. 164-167

Lleras-Muney A. and Shertzer A. (2015), *Did the americanization movement succeed? An evaluation of the effect of English-only and compulsory schooling laws on immigrants*, «American Economic Journal: Economic Policy», 7, 3, pp. 258-90

Lorenzo F.M.E., Galvez-Tan J. *et al.* (2007), *Nurse Migration from a Source Country Perspective: Philippine Country Case Study*, «Health Services Research», 42, 3-2, pp. 1406-1418

Machin S., Salvanes K.G. and Panu P. (2012), *Education and mobility*, «Journal of the European Economic Association», 10, 2, pp. 417-450

Malamud O. and Wozniak A.K. (2010a), *The impact of college education on migration: Evidence from the Vietnam generation*, «Journal of Human Resources», 47, 4 , pp. 913-950

– (2010b), *The impact of college education on geographic mobility: Identifying education using multiple components of Vietnam draft risk*, NBER wp n. 16463

McHenry P. (2013), *The relationship between schooling and migration: Evidence from compulsory schooling laws*, «Economics of Education Review», 35, pp. 24-40

McKenzie D. and Rapoport H. (2010), *Self-selection patterns in Mexico-US migration: The role of migration networks*, «Review of Economics and Statistics», 92, 4, pp. 811-821

– (2011), *Can migration reduce educational attainment? Evidence from Mexico*, «Journal of Population Economics», 24, pp. 1331-1358

McManus W.S. (1985), *Labor market costs of language disparity: An interpretation of Hispanic earnings differences*, «American Economic Review», 75, 4, pp. 818-827

McManus W.S., Gould W. and Welch F. (1983), *Earnings of Hispanic men: The role of English language proficiency*, «Journal of Labor Economics», 1, 2, pp. 101-130

Milligan K., Moretti E. and Oreopoulos P. (2004), *Does education improve citizenship? Evidence from the United States and the United Kingdom*, «Journal of Public Economics», 88, 9-10, pp. 1667-1695

Ministry of Education Macedonia (2004), *National Report for Development of Education in the Republic of Macedonia 2001-2004*

Molloy R., Christopher S.L. and Wozniak A. (2011), *Internal migration in the United States*, Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Aairs, Federal Reserve Board, Washington, D.C.

Mora M.T. and Davila A. (1998), *Gender, earnings, and the English skill acquisition of Hispanic workers in the United States*, «Economic Inquiry», 36, 4, pp. 631-644

Murtin F. and Viarengo M. (2011), *The Convergence Process of Compulsory Schooling in Western Europe: 1950-2000*, «Economica», 78, pp. 501-522

Nash R. and Eleftheriou M. (2008), *A review of the European schools language and science policies*, Marine Biodiversity & Ecosystem Functioning (MarBEF), EU Network of Excellence.

OECD (2001), *Reviews of National Policies for Education*: *Latvia*, OECD Centre for Co-operation with Non-Members, Organisation for Economic Co-operation and Development, Paris

– (2002), *Reviews of National Policies for Education: Lithuania*, OECD Centre for Co-operation with Non-Members, Organisation for Economic Co-operation and Development, Paris

Ortega F. and Peri G. (2009), *The causes and effects of international migrations: Evidence from OECD countries 1980-2005*, NBER wp n. 14833

Reiff P. (2012), *Le Luxembourg 1960-2010, L'évolution de l'enseignement depuis 1960*, Institut national de la statistique et des études economiques

Reinhold S. and Thom K. (2012), *Migration Experience and Earnings in the Mexican Labor Market*, «Journal of Human Resources», 48, 3, pp. 768-820

Roy A.D. (1951), *Some Thoughts on the Distribution of Earnings*, «Oxford Economic Papers», 3, 2, pp. 135-146

Saar E. 2008), *The Estonian educational system and the ISCED-97," in The International Standard Classication of Education (ISCED-97). An evaluation of content and criterion validity for 15 European Countries*, edited by S.L. Schneider, Mannheim, MZES

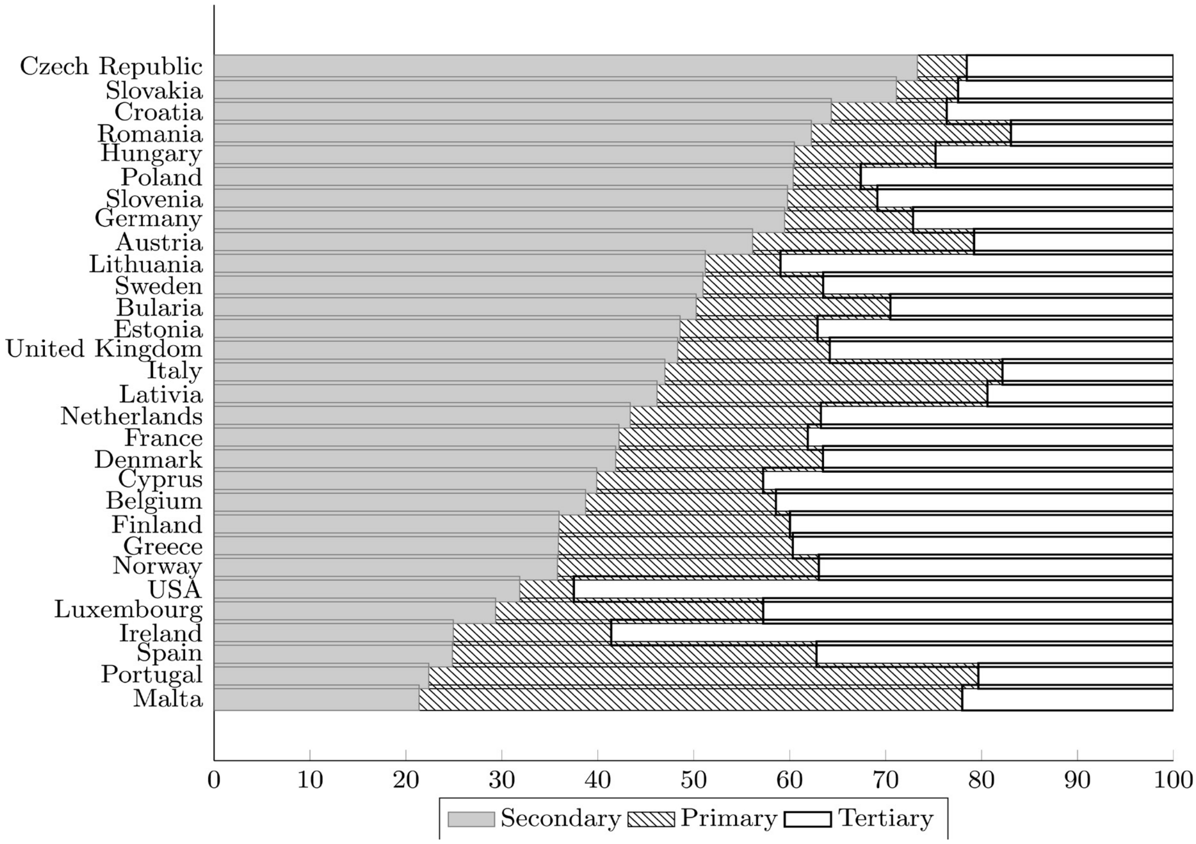
Stark O. (1991), *The migration of labor*, Basil Blackwell, Cambridge (MA), 1991

State Statistical Office, Republic of Macedonia (2015), *Primary, Lower and Upper Secondary Schools at the Beginning of the School Year 2014-2015*.

Tomich V. (1963), *Education in Yugoslavia and the New Reform The Legal Basis, Organization, Administration, and Program of, the Secondary Schools*, Office of Education, U.S. Department of Health, Education and Welfare

8. Appendix

Figure A-1 • Educational attainment of the population 25-44, 2010



Source: Barro and Lee (2010) primary: some secondary, primary and no schooling; secondary: secondary: secondary completed; tertiary: higher than secondary completed; Data for Liechtenstein not available.

TABLE A.1 • REFORMS: CHANGE IN LENGTH OF COMPULSORY EDUCATION

COUNTRY	YEARS		FIRST AFFECTED COHORT
	BEFORE	AFTER	
Belgium	8	12	1969
Bulgaria	8	9	1976
Czech Republic	9	10	1968
Czech Republic	10	9	1975
Estonia	8	11	1973
Estonia	11	9	1976
Finland	6	9	1970
Germany	9	13	1977
Latvia	8	11	1973
Latvia	11	9	1975
Lithuania	8	11	1973
Lithuania	11	9	1975
Luxembourg	9	10	1972
Malta	10	11	1983
Netherlands	10	11	1973
Netherlands	11	12	1980
Portugal	6	9	1980
Romania	10	8	1976
Slovakia	9	10	1968
Slovakia	10	9	1975
Slovakia	9	10	1984
Spain	8	10	1978

Sources: Brunello, Fort and Weber 2009; Hörner *et al*. 2007; National Education Act Bulgaria, Eurydice 1997; Saar 2008; Archimedes Foundation 2010; Garrouste 2010; OECD 2001; OECD 2002; Reiff 2012; Murtin and Viarengo 2011.

Table A.2 • Reforms: Foreign language classes in compulsory education

Origin country	Destination countries	First affected cohort
Austria	Ireland, UK	1975
Austria	Italy	1985
Belgium	Austria, Germany, Netherlands	1953
Belgium	Ireland, UK	1978
Bulgaria	Austria, Belgium, Germany, Ireland, UK	1974
Bulgaria	Italy, Spain	1982
Croatia	Austria, Germany, Ireland, UK	1948
Croatia	Italy	1971
Croatia	Belgium	1975
Cyprus	Ireland, UK	1951
Cyprus	Belgium	1960
Czech Republic	Austria, Belgium, Germany, Ireland, Spain, UK	1979
Denmark	Ireland, UK	1963
Estonia	Austria, Belgium, Germany, Ireland, UK	1983
Finland	Ireland, Sweden, UK	1961
Finland	Austria, Belgium, Germany, Italy, Spain	1985
France	Austria, Germany, Ireland, UK	1952
France	Spain, Italy	1985
Germany	Ireland, UK	1959
Greece	Belgium, Ireland, UK	1964
Greece	Austria, Germany	1979
Hungary	Austria, Germany, Ireland, UK	1979
Italy	Austria, Belgium, Germany, Ireland, Spain, UK	1952
Latvia	Austria, Belgium, Germany, Ireland, UK	1982
Liechtenstein	Austria, Belgium, Germany	1960
Liechtenstein	Ireland, UK	1988
Lithuania	Austria, Belgium, Germany, Ireland, UK	1980
Luxembourg	Austria, Germany	1962
Macedonia	Austria, Germany, Ireland, UK	1948
Macedonia	Italy	1971
Macedonia	Belgium	1975
Malta	Austria, Belgium, Germany, Italy	1993
Netherlands	Austria, Belgium, Germany, Ireland, UK	1951
Netherlands	Spain	1979
Netherlands	Italy	1987
Norway	Ireland, UK	1959
Poland	Austria, Belgium, Germany, Ireland, UK	1979

continues

follows

Portugal	Austria, Belgium, Germany, Ireland, UK	1976
Portugal	Spain	1987
Romania	Austria, Belgium, Germany, Ireland, UK	1957
Romania	Italy, Spain	1979
Slovakia	Austria, Belgium, Germany, Ireland, Spain, UK	1979
Slovakia	Italy	1983
Slovenia	Austria, Germany, Ireland, UK	1948
Slovenia	Italy	1971
Spain	Ireland, UK	1982
Sweden	Ireland, UK	1952
Sweden	Austria, Belgium, Germany, Spain	1981
UK	Austria, Belgium, Germany, Spain	1977

Sources: Directorate General for Education and Culture 2001; Braham 1972; Galvez *et al*. 2000; Education, Audiovisual and Culture Executive Agency 2012; Education, Audiovisual and Culture Executive Agency 2010; Ministry of Education Macedonia 2004; State Statistical Oce, Republic of Macedonia 2015; Nash and Eleftheriou 2008; Tomich 1963.

Table A.3 • Average years of education and length of compulsory education

Years of compulsory schooling	(1)	(2)	(3)
	0.331	0.348	0.314
	*(0.101)****	*(0.101)****	*(0.136)****
Obs.	108	108	108
R^2	0.201	0.225	0.747
F statistic	5.224	3.859	14.067

The dependent variable is the average number of years of education by cohort. The variable years of compulsory schooling refers to the average number of years of compulsory schooling faced by the corresponding cohort. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year-fixed effects, age indicators, binary variables for each pair of origin and destination countries, dummies for each combination of age and year, a variable for differences in lagged unemployment rate between origin and destination countries and the stock of co-nationals from each cohort in the destination country in the previous period. Errors are clustered by origin-destination-age.

Table A.4 • Robustness check: Migration and years of compulsory schooling also in destination country

Years of compulsory schooling	(1)	(2)	(3)	(4)
	-22.832	-43.134	-32.750	-31.295
	(15.549)	*(15.558)***	*(14.338)****	*(14.497)***
Years of compulsory schooling in destination	60.003	-3.645	-11.059	1547.559
	*(19.841)****	*(121.279)*	*(26.155)*	*(2535.695)*
Obs.	11,205	11,205	11,205	11,205
R^2	0.684	0.725	0.928	0.93

The dependent variable is the number of immigrants. The variable years of compulsory schooling refers to the average number of years of compulsory schooling faced by the corresponding cohort. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year-fixed effects, age indicators, binary variables for each pair of origin and destination countries, dummies for each combination of age and year, a variable for differences in lagged age-specific unemployment rate between origin and destination countries, the stock of co-nationals from each age group in the destination country in the previous period, and the size of the age group in the origin country. Errors are clustered by origin-destination-age.

Table A.5 • Robustness check: Migration and compulsory foreign language classes, excluding United Kingdom

Foreign language classes	(1)	(2)	(3)	(4)
	439.348	411.356	338.933	337.388
	*(137.011)****	*(163.869)***	*(153.727)***	*(156.849)***
Obs.	10,921	10,921	10,921	10,921
R^2	0.706	0.74	0.957	0.958

The dependent variable is the number of immigrants. The variable years of compulsory schooling refers to the average number of years of compulsory schooling faced by the corresponding cohort. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year-fixed effects, age indicators, binary variables for each pair of origin and destination countries, dummies for each combination of age and year, a variable for differences in lagged age-specific unemployment rate between origin and destination countries, the stock of co-nationals from each age group in the destination country in the previous period, and the size of the age group in the origin country. Errors are clustered by origin-destination-age.

TABLE A.6 • ROBUSTNESS CHECK: MIGRATION AND YEARS OF COMPULSORY SCHOOLING, INCL. YEARS LIVED UNDER COMMUNIST RULE

Years of compulsory schooling	(1)	(2)	(3)	(4)
	-24.618	-31.383	-32.706	-31.295
	(15.301)	*(15.238)***	*(14.346)***	*(14.497)***
Obs.	11,205	11,205	11,205	11,205
R^2	0.683	0.713	0.928	0.93

The dependent variable is the number of immigrants. The variable years of compulsory schooling refers to the average number of years of compulsory schooling faced by the corresponding cohort. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year-fixed effects, age indicators, binary variables for each pair of origin and destination countries, dummies for each combination of age and year, a variable for differences in lagged age-specific unemployment rate between origin and destination countries, the stock of co-nationals from each age group in the destination country in the previous period, and the size of the age group in the origin country. Errors are clustered by origin-destination-age.

TABLE A.7 • ROBUSTNESS CHECK: MIGRATION AND COMPULSORY FOREIGN LANGUAGE CLASSES, INCL. YEARS LIVED UNDER COMMUNIST RULE

Foreign language classes	(1)	(2)	(3)	(4)
	560.060	418.459	393.544	391.515
	*(150.675)****	*(161.834)****	*(161.491)***	*(165.454)***
Obs.	11,205	11,205	11,205	11,205
R^2	0.688	0.728	0.93	0.932

The dependent variable is the number of immigrants. The variable years of compulsory schooling refers to the average number of years of compulsory schooling faced by the corresponding cohort. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year-fixed effects, age indicators, binary variables for each pair of origin and destination countries, dummies for each combination of age and year, a variable for differences in lagged age-specific unemployment rate between origin and destination countries, the stock of co-nationals from each age group in the destination country in the previous period, and the size of the age group in the origin country. Errors are clustered by origin-destination-age.

TABLE A.8 • ROBUSTNESS CHECK: EXOGENEITY OF CHANGES IN LENGTH OF COMPULSORY SCHOOLING

	population	geography	education	agriculture	manufacturing	economiccycle
	(1)	(2)	(3)	(4)	(5)	(6)
Population growth (annual %)	0.002 (0.009)	0.002 (0.009)	0.001 (0.01)	0.025 (0.035)	0.028 (0.035)	0.014 (0.037)
Rural population (% total pop.)		0.00004 (0.002)	0.0001 (0.002)	0.003 (0.005)	0.003 (0.005)	0.005 (0.005)
Average years of schooling			-.005 (0.013)	-.011 (0.028)	-.010 (0.028)	-.008 (0.028)
Value-added agriculture (%)				-.010 (0.014)	-.007 (0.015)	-.011 (0.016)
Value-added manufacturing (%)					0.004 (0.008)	0.005 (0.008)
GDP per capita						4.66e-06 (3.66e-06)
Unemployment rate						
Obs.	1,448	1,448	1,398	460	460	460
R^2	0.056	0.056	0.057	0.132	0.132	0.136
F statistic	1.047	1.033	1.009	0.953	0.941	0.953

The dependent variable is a dummy variable that takes on value one if a reform regarding the length of compulsory schooling was passed that year. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year and country of origin-fixed effects.
Sources: World Bank Data for population growth, percentage of rural population, value-added shares manufacturing and agriculture; Barro and Lee 2010 for average years of schooling; OECD Data for GDP per capita and unemployment rates.

TABLE A.9 • ROBUSTNESS CHECK: EXOGENEITY OF FOREIGN LANGUAGE CLASSES DURING COMPULSORY EDUCATION

	population	geography	education	agriculture	manufacturing	economic cycle
	(1)	(2)	(3)	(4)	(5)	(6)
Population growth (annual %)	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)	0.009 (0.009)	0.009 (0.009)	0.01 (0.009)
Rural population (% total pop.)		0.001 (0.0004)***	0.001 (0.0005)**	-.0002 (0.001)	-.0001 (0.001)	-.0004 (0.001)
Average years of schooling			-.001 (0.004)	0.004 (0.007)	0.003 (0.007)	0.003 (0.007)
Value-added agriculture (%)				-.0003 (0.004)	-.0007 (0.004)	-.0002 (0.004)
Value-added manufacturing (%)					-.0006 (0.002)	-.0007 (0.002)
GDP per capita						-5.82e-07 (9.05e-07)
Unemployment rate						
Obs.	1,448	1,448	1,398	460	460	460
R^2	0.048	0.051	0.052	0.133	0.133	0.134
F statistic	0.885	0.933	0.921	0.961	0.945	0.935

The dependent variable is a dummy variable that takes on value one if a reform regarding the length of compulsory schooling was passed that year. The coefficients are marked with * if the level of significance is between 5 and 10 per cent, ** if the level of significance is between 1 and 5 per cent and *** if the level of significance is less than 1 per cent. All regressions contain year and country of origin-fixed effects.
Sources: World Bank Data for population growth, percentage of rural population, value-added shares manufacturing and agriculture; Barro and Lee 2010 for average years of schooling; OECD Data for GDP per capita and unemployment rates.

Xingna Nina Zhang

Analysis of urban-rural divide and interprovincial migration in China from 2000 to 2010 with gravity models

Abstract. This paper investigates how China's interprovincial migration indicators changed between 2000 and 2010, and compares 2010's four different categories of migration streams (urban-rural, urban-urban, rural-urban and rural-rural), using censuses and gravity models and focusing on rural-urban divide. The results show that the (average) effect of rural/urban segment populations are remarkably significant and responsive to interprovincial migration, and that income is an increasingly significant factor, while distance gradually falls in prominence between 2000 and 2010. As with 2010's four migration streams, besides the constant deterrent impact of distance across all the models, the results confirm that destination urban population and income each has bigger impacts than their origin counterparts overall, while destination rural population and income are less significant than their origin counterparts respectively, proving that the pull force plays a bigger role than push force for urban factors. These findings are closely related to the unbalanced regional development in China, where the rural-urban divide has remained pronounced for a few decades. To conclude, rural/urban segment populations and incomes are realistic measures of directional and network-induced interprovincial migration within China, and the attraction from urban areas increases over time with income's rising significance while both pull and push forces are getting weaker in rural areas from 2000 to 2010, which mirrors the dynamics of interprovincial migration.

Keywords. China, interprovincial migration, gravity models, census data, rural-urban divide

1. Introduction

Internal migration has been an important topic in population studies, and many efforts have been made to quantitatively estimate it with different approaches and models in China's context (Huang, Li, Li *et al.* 2015; Jiang, Wang, Le *et al.* 2013; Li, Liu and Tang 2014; Shen 2015; Shi, Zheng, Sun *et. al.* 2014; Song and Liu 2014; Sun, Wang and Bai 2014; Wang, Chen and Li 2013; Yang, Han and Song 2014). These models basically fall into two categories: one group seeks to identify the motivational factors underpinning population movement, and to provide insights into migration's mechanism; the other attempts to simulate or project migration flows based on current or historical trends, aiming to uncover their spatiotemporal patterns (Stillwell 2005; Stillwell and Congdon 1991). Both strands are important (Kleinwechter and Grethe 2015; Mai, Peng, Dixon *et. al.* 2014), but a brief review of the former only is provided here as it is closely relevant to this paper.

In modelling China's internal migration mechanism, many researchers focus on identifying its contributing factors (Liu, Xie, Zhang *et. al.* 2014; Song and Liu 2014; Sun, Wang and Bai 2014; Wang, Chen and Li 2013). One common theme of these studies addresses the significance of rural migrants, and a lot of work has been done to utilise census data to look into internal migration on the whole (Bao, Shi and Hou 2005; Cai and Wang 2003a; 2003b; Chan 2011; Li 2009; Pang 2001; Song and Wang 2005). However, there is still a lack of investigation directly relating migration flows, rural-urban migration for instance, to motivational factors such as rural-urban divide, which leads to a limiting effect on the explanatory analysis.

The lack of proper exploration in this aspect, especially the rural-urban divide, is quite conspicuous in the current literature, despite the sound theoretical basis of the rural and urban classification of the space and economy (Wu and Yao 2003). Although there are many rural-urban migration studies around the world (Christian and Braden 1966; Claeson 1969; Fitzgerald, Leblang and Teets 2014; Flores, Zey and Hoque 2013; Ginsberg 1972; Johnston 1970; Peeters 2012; Poprawe 2015), few explicit investigations about the role of rural-urban divide in China's internal migration system have been made, while the majority are based on the unfounded assumption of homogeneous interprovincial migration flows.

Consequently, the ongoing task of recognising the inherent heterogeneity within interprovincial migrants based on their rural or urban origins and destinations as well as identifying determinants of different migration streams, still remains unfulfilled in studying China's interprovincial migration flows. This is the challenge that this paper has taken up by using different measures of independent and dependent migration variables with gravity models, drawing on knowledge and lessons from former studies.

The reason to choose the gravity model lies in its ability in incorporating origin and destination factors when migration itself is directly modelled, as it could be combined well with different model specifications and estimation methods (Fitzgerald, Leblang and Teets 2014; Flores, Zey and Hoque 2013; Peeters 2012; Poprawe 2015; Shen 2015). Although only distance and populations are used in its general form, its extended forms have much potential in explaining migration by adding in other socio-economic factors (Christian and Braden 1966; Claeson 1969; Ginsberg 1972; Johnston 1970). In fact, it is widely agreed as a good tool and the most popular model in describing and analysing migration flows (Fan 2005).

Gravity models could also find good grips from migration stock and push-pull theories. The former refers to the propensity of migrants to go to specific destinations is expected to be negatively correlated with distance (Fan 2005), while the latter bases its hypothesis on the positive relation between migration magnitude and origin or destination populations. Namely, apart from spatial frictions, a specific migrant flow receives the push from the origin and the pull from the destination at the same time (Diamantides 1992a, 1992b; Dorigo and Tobler 1983; Fan 1996; Heisler 1992; Mohtadi 1990).

However, studies using it to directly model China's internal migration flows are relatively scarce (Fan 2005). The majority of previous studies have focused on the exploration of migration determinants (Li, Liu and Tang 2014; Liu, Stillwell, Shen *et. al.* 2014; Shen 2015), its influences on sending or receiving areas (Song and Liu 2014; Sun, Wang and Bai 2014), spatial patterns (Fan 2005; Liu and Shen 2014; Wang, Chen and Li 2013; Yang, Han and Song 2014), and future trends (Jiang, Wang, Le *et. al.* 2013). In the few studies directly modelling

China's internal migration with gravity model, total population and GDP (or GDP *per capita*) are extensively used (Fan 2005; Mi, Zhou and Shi 2009; Shen 2015; Wang 1993), which might be imperfect measures of migration determinants and could be replaced with rural/urban segment populations and incomes. Two reasons account for this: first, it is impossible to estimate internal migration in China without taking into account of urban-rural divide; second, China's internal migration is inherently rural/urban segmented, with rural migrants being a major part of total migration population (for example, rural migrants account for 74.21 per cent of the total in 2010's census).

Similar to international migration, a variety of socio-economic variables play a key role in observed patterns of internal migration (Fitzgerald, Leblang and Teets 2014). But internal migration seems to more obey Zelinsky's mobility transition theory, which emphases the effects of urban-rural divide (Gedik 2005). As a country with vast rural population and urban-rural divide, China has been experiencing huge internal migration for the past few decades. Thus the fact that little research has studied the effects of rural-urban divide upon interprovincial migration on the national scale could be detrimental, as the joint relationship between population and income across provinces can bias estimates and unobserved variation could further complicate the separate identification of across province interactions. This makes it impossible to investigate the full picture of China's internal migration.

In view of this, this paper adopts a different approach to estimate the contribution of segmented rural/urban populations and incomes in China's interprovincial migration for the year of 2000 and 2010, by taking a step further to validate their relationships based on former gravity model studies (Mi, Zhou and Shi 2009; Shen 2015; Wang 1993; Wang, Chen and Li 2013). Additionally, this paper also extends the investigation to different interprovincial migration streams in 2010, aiming to uncover their unique mechanisms, which is the first ever attempt in this field.

This paper's structure is as follows: the next section provides a description of data, followed by a detailed explanation of gravity models and results, ending with discussion of key findings and conclusion of possible future research plan in studying China's internal migration.

2. Data

All the variables are listed in Table 1, drawn from Census and China Statistical Yearbook in the year of 2000 and 2010.

Censuses have 2 datasets, long-form (10 per cent sampling) and short-form (whole population), and the unit difference between them has been eliminated in this study. Most population data are directly from short-form dataset, with the exception of urban-rural, rural-urban, rural-rural and urban-urban migration streams. Instead, they are calculated based on detailed rural and urban origin and destination information in 2010's long-form dataset, where «Street» and «Neighbourhood committees of the town» are recognised as urban areas and «Township» and «Village committees of the town» as rural areas. But no such information is available in either 2000's long-form or short-form dataset.

In total, 961 migration streams are formed between 31 provinces for each year, among which 930 are interprovincial and used to analyse the longitudinal change from 2000 to 2010.

Meanwhile, 897 inter-provincial migration streams, with detailed rural/urban origin and destination information, are used to analyse their relationships in 2010.

Income data are from yearbooks. With temporal and rural/urban specific deflators, the average *per capita* income for rural/urban households is re-measured to consolidate the comparison between 2001 and 2011 across time and space. Additionally, origin-destination distance is calculated by the distance between provincial capitals.

All the variables are taken Euler Number logs before they are modelled, in order to facilitate the conduct of linear regression, which will be further explained in the next section.

TABLE 1 • VARIABLE LIST

Dependent variables	Units	Independent variables	Units
Interprovincial migration	000s persons	Origin-destination distance	000s m
Urban-rural migration	00s persons	Origin urban population	000,000s persons
Rural-urban migration	00s persons	Origin rural population	000,000s persons
Rural-rural migration	00s persons	Origin urban income	000s yuan
Urban-urban migration	00s persons	Origin rural income	000s yuan
		Destination urban population	000,000s persons
		Destination rural population	000,000s persons
		Destination urban income	000s yuan
		Destination rural income	000s yuan

3. METHODS

In modelling migration, the original gravity model could be written as

$$m_{ij} = \mathrm{k}\frac{p_i p_j}{d_{ij}^2} \tag{1}$$

or the more general form

$$m_{ij} = \mathrm{k} \times p_i^a \times p_j^b \times d_{ij}^c \tag{2}$$

where

- a, b and c are parameters;
- m_{ij} is migration flow between place i and place j;
- p_i is the population in origin place i;
- p_j is the population in destination place j;
- d_{ij} is the distance between place i and place j.

If incomes I_i and I_j of origin i and destination j are added in, plus their parameters g and f, the equation becomes

$$m_{ij} = \mathrm{k} \times p_i^a \times p_j^b \times {I_i}^f \times {I_j}^g \times d_{ij}^c \tag{3}$$

If taken logs on both sides, Equation 3 becomes

$$log\, m_{ij} = logk + \mathrm{a}\times log\, p_i + \mathrm{b}\times log\, p_j \tag{4}$$

The main reason for using log-linear modelling is that it can provide information on main and interaction effects as part of the same analysis (Flores, Zey and Hoque 2013), and here both linear and multi-variate linear regressions are used.

Based on push-pull and migration stock theories, the decomposition of population and income in Equation 4 is essential in order to investigate the effects of rural-urban divide, and it becomes

$$\begin{aligned} log\, m_{ij} = logk &+ \mathrm{a}_1\times log\, p_{ir} + \mathrm{a}_2\times log\, p_{iu} + \mathrm{b}_1\times log\, p_{jr} + \mathrm{b}_2\times log\, p_{ju} \\ &+ \mathrm{c}\times log\, d_{ij} + \mathrm{f}_1\times log\, I_{ir} + \mathrm{f}_2\times log\, I_{iu} + \mathrm{g}_1\times log\, I_{jr} \\ &+ \mathrm{g}_2\times log\, I_{ju} \end{aligned} \tag{5}$$

$$\begin{aligned} M_{2000} = 21.117\times {p_{or}}^{0.943}\times {p_{ou}}^{0.166}\times {p_{dr}}^{0.017}\times {p_{du}}^{0.636}\times {I_{or}}^{0.492} \\ \times {I_{ou}}^{-1.118}\times {I_{dr}}^{0.155}\times {I_{du}}^{2.002}\times d^{-1.096} \end{aligned} \tag{6}$$

where

- k, a_1, a_2, b_1, b_2, c, f_1, f_2, g_1 *and* g_2 are parameters;
- p_{ir} and p_{iu} are the rural and urban population of origin province i respectively;
- p_{jr} and p_{ju} are the rural and urban population of destination province j respectively;
- I_{ir} and I_{iu} are the rural and urban household income *per capita* of origin province i respectively;
- I_{jr} and I_{ju} are the rural and urban household income *per capita* of destination province j respectively.

Thus, the extended and enhanced gravity model of migration is established, where distance measures the friction in space while rural and urban segment incomes and populations measure economic and migration stock impacts respectively, with the underlying hypothesis that China's internal migration is so highly directional and social network-induced that rural and urban segment incomes and populations are more responsive to it than their total counterparts.

4. Results

4.1 Interprovincial migration in 2000 and 2010

This section features exploring 2000 and 2010's comparison with linear regression models. First, models without taking logs on variables are run with both years' data, results confirming that they could not explain interprovincial migration as well as its log-linear counterparts with adjusted R^2 being 0.143 and 0.192 (Table 6 of the Appendix), but rising to 0.653 (Model 1) and 0.737 (Model 2) for 2000 and 2010 respectively after taking logs.

$$M_{2010} = 1.725 \times p_{or}{}^{0.669} \times p_{ou}{}^{0.480} \times p_{dr}{}^{-0.330} \times p_{du}{}^{0.942} \times I_{or}{}^{0.104} \times I_{ou}{}^{-1.715} \times I_{dr}{}^{-1.437} \times I_{du}{}^{4.327} \times d^{-1.002} \quad (7)$$

Thus the choice of log-linear regression models is consolidated, with results presented in Table 2. Both years' models predict interprovincial migration significantly well, with most coefficients being significant. Coefficients of origin rural and destination urban populations are significant and positive in both models, while destination rural population is significant and negative in 2010. About incomes, more variations are shown: coefficients of destination urban income are always significant and positive while origin urban income remains to be significantly negative in both years; coefficients of destination rural income change from being insignificantly positive to significantly negative from 2000 to 2010. Additionally, both coefficients of distance are significant and negative, proving its adverse effects in interprovincial migration. When transformed back to gravity models, their equations (6 and 7) could be written as above, with changes in coefficients presented in Table 3.

TABLE 2 • INTERPROVINCIAL MIGRATION MODEL COEFFICIENTS

	2000		2010	
Parameter	**B**	**S. E.**	**B**	**S. E.**
Origin urban population	0.166	0.113	0.480***	0.091
Origin rural population	0.943***	0.083	0.669***	0.077
Destination urban population	0.636***	0.113	0.942***	0.091
Destination rural population	0.017	0.083	-0.330***	0.077
Distance	-1.096***	0.067	-1.002***	0.053
Origin urban income	-1.118**	0.322	-1.715***	0.377
Origin rural income	0.492	0.272	0.104	0.289
Destination urban income	2.002***	0.322	4.327***	0.377
Destination rural income	0.155	0.272	-1.437***	0.289
Constant	3.05	0.912	0.545	1.052

Note: all the parameters are from Table 2, so please refer to the earlier table for their coefficients and standard errors. *** denotes $p<0.001$, ** represents $p<0.01$, and * symbolises $p<0.05$.

Here,

- M_{2000} and M_{2010} are interprovincial migration flows in 2000 and 2010 respectively;
- all the independent variables follow definitions in Equation 5 but in according years.

In 2000, for every time's increase in each independent variable, interprovincial migration's responses are detailed by holding all other variables constant in turn: for origin and destination rural and urban populations, 1.923, 1.122, 1.012 and 1.554 times' changes are predicted; for origin and destination urban and rural incomes, 0.461, 1.406, 4.006 and 1.113

time(s)' changes are expected; and a 0.468 time's change for distance. Similarly, 2010's interprovincial migration's responses are: 1.395, 1.590, 1.921 and 0.796 time(s)' change for origin and destination urban and rural populations, 0.305, 1.075, 20.070 and 0.369 time(s)' change for origin and destination urban and rural incomes, and 0.499 time's change for distance.

TABLE 3 • INTERPROVINCIAL MIGRATION'S RESPONSES FOR EVERY TIME'S INCREASE IN INDEPENDENT VARIABLES

	Changes in dependent variable	
	2000	2010
Origin urban population	1.122	1.395
Origin rural population	1.923	1.590
Destination urban population	1.554	1.921
Destination rural population	1.012	0.796
Distance	0.468	0.499
Origin urban income	0.461	0.305
Origin rural income	1.406	1.075
Destination urban income	4.006	20.070
Destination rural income	1.113	0.369

Note: all the parameters are from Table 2. Please refer to it for their coefficients and standard errors.

As independent variables might have gone through significant changes within 10 years, directly comparing interprovincial migration models between 2000 and 2010 could be difficult. However, provincial capital distances changed little, and rural populations remained almost the same with a 0.6 per cent national increase. That is to say, the reason that interprovincial migrants are more than doubled between 2000 and 2010 lies in other dependent variables' changes, with urban population, and rural and urban income each experiencing a 26.0, 162.7 and 204.3 per cent growth respectively. In summary, distance and origin and destination rural population gradually fall in significance, origin and destination urban populations and incomes rise in prominence by comparison, while origin and destination rural incomes decrease in the estimated parameter values despite their remarkable growth from 2000 to 2010.

4.2 Interprovincial migration streams in 2010

In this section, 2010's four specific interprovincial migration streams, namely rural-urban, rural-rural, urban-rural and urban-urban, is extensively studied with multi-variate linear regression (Table 4), with each accounting for 44.24, 29.97, 3.42 and 22.38 per cent of the total and evidencing rural-urban stream's predominance in 2010. It is also logically valid to compare with 2010's interprovincial model (Table 5), as they all use the same set of variables.

TABLE 4 • INTERPROVINCIAL MIGRATION STREAM MODEL COEFFICIENTS

	Urban-urban migration		Urban-rural migration		Rural-urban migration		Rural-rural migration	
Parameter	B	S.E.	B	S.E.	B	S.E.	B	S.E.
Origin urban population	0.635***	0.079	0.804***	0.104	-0.071	0.116	0.316*	0.134
Origin rural population	0.16*	0.064	0.12	0.085	1.361***	0.094	0.954***	0.108
Destination urban population	1.037***	0.079	0.249*	0.104	0.687***	0.115	0.586***	0.133
Destination rural population	-0.438***	0.064	0.213*	0.084	-0.219*	0.093	0.036	0.108
Distance	-0.854***	0.043	-1.045***	0.057	-0.991***	0.063	-1.302***	0.073
Origin urban income	-2.258***	0.313	-1.491***	0.413	-3.535***	0.457	-0.951	0.528
Origin rural income	1.119***	0.24	-0.067	0.316	1.012**	0.35	-1.046**	0.404
Destination urban income	3.738***	0.312	2.417***	0.41	4.455***	0.455	3.457***	0.525
Destination rural income	-1.42***	0.24	-0.89**	0.316	-1.084**	0.35	-0.938**	0.405
Constant	5.586***	0.869	6.749***	1.145	6.049***	1.268	5.395***	1.466

In urban-urban migration equation 8 (Model 3, Adjusted $R^2 = 0.718$), for every time's increase in each independent variable, urban-urban stream's responses are presented by holding all other variables constant in turn: for origin/destination urban/rural population, 1.553, 1.117, 2.052 and 0.738 time(s)' changes are predicted; for origin/destination urban/rural incomes, 0.209, 2.172, 13.343 and 0.374 time(s)' changes are expected; a 0.553 time's change for distance. Similarly, urban-rural stream's responses are (Model 4, Adjusted $R^2 = 0.566$): 1.746, 1.087, 1.188 and 1.159 time(s)' change for origin/destination urban/rural population, 0.356, 0.955, 0.540 and 5.341 for origin/destination urban/rural incomes, and 0.485 for distance.

$$M_{uu} = 2.666 \times {p_{or}}^{0.160} \times {p_{ou}}^{0.635} \times {p_{dr}}^{-0.438} \times {p_{du}}^{1.037} \times I_{ou}^{-2.258} \times I_{or}^{1.119} \times I_{du}^{3.738} \times I_{dr}^{-1.420} \times d^{-0.854} \tag{8}$$

$$M_{ur} = 8.534 \times {p_{or}}^{0.120} \times {p_{ou}}^{0.804} \times {p_{dr}}^{0.213} \times {p_{du}}^{0.249} \times I_{ou}^{-1.491} \times I_{or}^{-0.067} \times I_{du}^{2.417} \times I_{dr}^{-0.890} \times d^{-1.045} \tag{9}$$

Where,

- M_{uu} and M_{ur} are the number of migrants from origin province's urban to destination province's urban and rural areas respectively.

In rural-urban migration equation 10 (Model 5, Adjusted R^2 =0.711), for every time's increase in each independent variable, rural-urban stream's responses are presented by holding all other variables constant in turn: for origin/destination urban/rural population, 0.952, 2.569, 1.610 and 0.859 time(s)' changes are predicted; for origin/destination urban/rural incomes, 0.086, 2.017, 21.933 and 0.472 time(s)' changes are expected; a 0.503 time's change for distance. Similarly, rural-rural stream's responses are (Model 6, Adjusted R^2 = 0.634): 1.245, 1.937, 1.501 and 1.025 time(s)' change for origin/destination urban/rural population, 0.517, 0.484, 10.981 and 0.522 for origin/destination urban/rural incomes, and 0.406 for distance.

$$M_{ru} = 4.236 \times {p_{or}}^{1.361} \times {p_{ou}}^{-0.071} \times {p_{dr}}^{-0.219} \times {p_{du}}^{0.687} \times I_{ou}^{-3.535} \times I_{or}^{1.012} \times I_{du}^{4.455} \times I_{dr}^{-1.084} \times d^{-0.991} \qquad (10)$$

$$M_{rr} = 2.202 \times {p_{or}}^{0.954} \times {p_{ou}}^{0.316} \times {p_{dr}}^{0.036} \times {p_{du}}^{0.586} \times I_{ou}^{-0.951} \times I_{or}^{-1.046} \times I_{du}^{3.457} \times I_{dr}^{-0.938} \times d^{-1.302} \qquad (11)$$

Where,

- M_{ru} and M_{rr} each is the number of migrants from origin province's rural to destination province's urban and rural areas.

TABLE 5 • COMPARISON OF DEPENDENT VARIABLES' RESPONSES FOR EVERY TIME'S INCREASE IN EACH INDEPENDENT VARIABLE

	Responses in dependent variable				
	Model 3	Model 4	Model 5	Model 6	Model 2
Origin urban population	1.553	1.746	0.952	1.245	1.395
Origin rural population	1.117	1.087	2.569	1.937	1.590
Destination urban population	2.052	1.188	1.610	1.501	1.921
Destination rural population	0.738	1.159	0.859	1.025	0.796
Distance	0.553	0.485	0.503	0.406	0.499
Origin urban income	0.209	0.356	0.086	0.517	0.305
Origin rural income	2.172	0.955	2.017	0.484	1.075
Destination urban income	13.343	5.341	21.933	10.981	20.070
Destination rural income	0.374	0.540	0.472	0.522	0.369

Note: all the parameters are from Table 4. Please refer to it for their coefficients and standard errors.

As shown by Table 5, Model 2 describes 2010's total interprovincial migration, while Model 3, 4, 5 and 6 are its subsets with specific urban and rural destinations and origins. Thus, Model 2 could act as the baseline when conducting comparisons: Model 3 has smaller rural but bigger urban population impacts, and bigger rural but smaller urban income effects in both origin and destination, with evidently the smallest adverse effect of distance; Model 4 has bigger origin urban and destination rural population and income effects but smaller origin rural and destination urban population and income impacts at the same time, with bigger distance adverse effect; Model 5 has bigger rural population and income impacts but smaller urban population effects in both origin and destination, with the biggest destination but smallest origin impacts as for urban income and the second smallest distance adverse effect; Model 6 has bigger rural but smaller urban population impacts in both origin and destination, and bigger origin urban and destination rural but smaller origin rural and destination urban income effects, with the largest distance adverse influence.

Thus, several important messages are drawn: destination urban income plays a predominant facilitating role, while origin urban and destination rural incomes act as obstructers for all the streams; origin rural income acts as the facilitator in streams of urban destination but the obstructer in streams of rural destination simultaneously; although distance always remains as the deterrent, rural-destination streams have bigger distance decay compared with their urban-destination counterparts; destination urban and origin rural populations are facilitators for all the streams, and so is origin urban population except for rural-urban stream, while destination rural population acts as the stimulus in rural-destination streams but the deterrent in urban-destination streams. Overall, destination urban population and income seems to be more significant than their origin counterparts, so does origin rural population and income compared with their destination counterparts, suggesting that the pull force plays a bigger role than push force for urban factors, and that bigger push force exists as for rural factors.

5. Discussion

From 2000 to 2010, interprovincial migration is more than doubled in volume, with results above confirming: as distance and (both origin and destination) rural populations fall in prominence, spatial friction decreases and rural populations provide less facilitating over time, probably owing to infrastructure improvement and countryside's decline; interprovincial migration is shifting to urban areas as the centre for both sending and receiving interprovincial migrants, as urban populations and incomes rise while rural incomes decrease in prominence at both origins and destinations. During this period, China's economy and society experienced great improvement (Lin, Cai and Li 2003; Luo, Shen and Gu 2014), accompanied by a declining origin but increasing destination effect on the whole. Thus, it is proved that income facilitates the production of migrants and economy is getting increasingly important in relocating populations, and that the tide is changing as destination economy is gaining force in attracting migrants.

As with 2010's four migration streams, the results confirm that the pull force plays a bigger role than push force for urban factors, and that bigger push force exists for rural factors in comparison, indicating that urban areas are in the predominant position among all the four migration types. This is closely related to the unbalanced regional development in China, where the urban-rural divide has remained to be pronounced for the past few decades and rural areas are usually rated as the sending origin while their urban counterparts receiving destinations.

In summary, the attraction from urban areas increases over time from 2000 to 2010 accompanying booming economy and society while both pull and push forces are getting weaker in rural areas, the (average) effect of rural and urban segment population and income is markedly significant in 2010. As economic factors get increasingly significant, it is facilitating the mobilization of resources, humans (labour or human capital) included, and with technology is redefining «distance» to a large extend to make space ever smoother for the mobilization and relocation of resources. In other words, places have to take on greater responsibility for strategizing and implementing economic development activities to produce attraction to «stick» those migrating resources and make them settle down and get involved with local development.

6. Conclusion

This paper's empirical analysis has shown that the extended and enhanced gravity model is indeed effective to describe and explain interprovincial migration from 2000 to 2010 and its four categories of streams in 2010 in China, which also helps to explain migrant's directional movement and what factors affect their destination choices. In this sense, this paper enhances both the theoretical and empirical understanding of China's internal migration.

Certainly, internal population movement is not only unique to China, but its magnitude and massiveness is stunning even in the whole human history. China's case is also special in that its fast urbanization and industrialization process has made the whole nation hastily plunge into rapid population relocation, with lingering socio-economic and institutional barriers inflaming rural migrants' inferiority and social issues in rural China such as brain-drain and people left-behind.

Rather, more unknown could be explored in relation to the socio-economic transitions that China have been experiencing for decades. Future research is needed to reveal how much the change of migration patterns is due to demographic and socio-economic indicators, and fruitful work could also be done in improving estimation precision by combining multilevel modelling to investigate origin/destination effects.

References

Bao S., Shi A. and Hou W. (2005), *Zhongguo renkou qianyi de kongjian xingtai bianhua fenxi*, «Chinese Journal of Population Science», 2005, 5, pp. 28-36

Cai F. and Wang D. (2003a), *Migration as marketization: what can we learn from China's 2000 census data?*, «China Review», 3, 2, pp. 73-93

- (2003b), *Zuowei shichanghua de renkou liudong diwuci renkou pucha shuju fenxi*, «Zhongguo renkou kexue», 2003, 5, pp. 11-19

Chan K.W. (2011), *Internal migration in China: Trends, geography and policies*. «Population distribution, urbanization, internal migration and development: An international perspective», pp. 81-109

Christian W. and Braden W. (1966), *Rural migration and gravity model*, «Rural Sociology», 31, 1, pp. 73-80

Claeson C.F. (1969), *2-Stage model of in-migration to urban centres: Deductive development of a variant of gravity formulation*, «Geografiska Annaler Series B-Human Geography», 51, 2, pp. 127-138

Diamantides N.D. (1992a), *The macrodynamics of international migration as a sociocultural diffusion process. A. Theory*, «Technological Forecasting and Social Change», 42, 3, pp. 273-289

- (1992b), *The macrodynamics of international migration as a sociocultural diffusion process. B. Applications*, «Technological Forecasting and Social Change», 42, 4, pp. 385-408

Dorigo G. and Tobler W. (1983), *Push-pull migration laws*, «Annals of the Association of American Geographers», 73, 1, pp. 1-17

Fan C.C. (1996), *Economic opportunities and internal migration: A case study of Guangdong Province, China*, «Professional Geographer», 48, 1, pp. 28-45

- (2005), *Modeling interprovincial migration in China, 1985-2000*, «Eurasian Geography and Economics», 46, 3, pp. 165-184

Fitzgerald J., Leblang D. and Teets J.C. (2014), *Defying the law of gravity. The political economy of international migration*, «World Politics», 66, 3, pp. 1-49

Flores M., Zey M. and Hoque N. (2013), *Economic liberalization and contemporary determinants of Mexico's internal migration: An application of spatial gravity models*, «Spatial Economic Analysis», 8, 2, pp. 195-214

Gedik A. (2005), *Mobility transition (test of Zelinsky's theory) and economic and demographic factors: japan and turkey, 1955-2000*, http://demoscope.ru/weekly/knigi/tours_2005/papers/iussp2005s51383.pdf

Ginsberg R.B. (1972), *Incorporating causal structure and exogenous information with probabilistic models - With special reference to choice, gravity, migration, and Markov Chains*, «Journal of Mathematical Sociology», 2, 1, pp. 83-103

Heisler B.S. (1992), *The future of immigrant incorporation - Which models - Which concepts*, «International Migration Review», 26, 2, pp. 623-645

Huang J.P., Li C., Li Q.Y. *et. al.* (2015), *Least-squares reverse time migration with static plane-wave encoding*, «Chinese Journal of Geophysics-Chinese Edition», 58, 6, pp. 2046-2056

Jiang C., Wang J., Le Y. *et al.* (2013), *Population spatial migration tendency forecasting with Grey model and Fourier series*, «Sustainable Development and Environment Ii, Pts 1 and 2», 409-410, pp. 69-74

Johnston R.J. (1970), *Latent migration potential and gravity model - New-Zealand study*, «Geographical Analysis», 2, 4, pp. 387-397

Kleinwechter U. and Grethe H. (2015), *National wage trends and migration in a Chinese village economy: A micro level modeling approach based on a composite utility function*, «American Journal of Agricultural Economics», 97, 3, pp. 701-726

Li P. (2009), *Zhongguo chengxiang renkou qianyi de shikong tezheng jiqi yingxiang yinsu*, «Economist», 1, pp. 50-57

Li Y., Liu H. and Tang Q. (2014), *Analysis of determinants on China's interprovincial migration during 1985-2005*, «Sustainable Development of Industry and Economy, Pts 1 and 2», 869-870, pp. 1096-1105

Lin J.Y., Cai F. and Li Z. (2003), *The China Miracle: Development Strategy and Economic Reform*, Chinese University Press, Hong Kong

Liu S.Q., Xie F.T., Zhang H.Q. *et al.* (2014), *Influences on rural migrant workers' selection of employment location in the mountainous and upland areas of Sichuan, China*, «Journal of Rural Studies», 33, pp. 71-81

Liu Y. and Shen J.F. (2014), *Spatial patterns and determinants of skilled internal migration in China, 2000-2005*, «Papers in Regional Science», 93, 4, pp. 749-771

Liu Y., Stillwell J., Shen J.F. *et al.* (2014), *Interprovincial migration, regional development and state policy in China, 1985-2010*, «Applied Spatial Analysis and Policy», 7, 1, pp. 47-70

Luo X., Shen J. and Gu C. (2014), *Urban and regional governance in China: Introduction*, «China Review», 14, 1, pp. 1-9

Mai Y.H., Peng X.J., Dixon P. *et al.* (2014), *The economic effects of facilitating the flow of rural workers to urban employment in China*, «Papers in Regional Science», 93, 3, pp. 619-642

Mi H., Zhou W. and Shi W. (2009), *Renkou qianyi zhongli moxing de xiuzheng ji qi yingyong*, «Population Study», 33, 4, pp. 99-104

Mohtadi H. (1990), *Rural-inequality and rural-push versus urban-pull migration - The case of Iran, 1956-1976*, «World Development», 18, 6, pp. 837-844

Pang L. (2001), *Duocengci fenxi fangfa zai renkou qianyi zhong de yingyong: shengji laodongli qianyi de duocengci fenxi*, «China Rural Survey», 2001, 2, pp. 11-17

Peeters L. (2012), *Gravity and spatial structure: The case of interstate migration in Mexico*, «Journal of Regional Science», 52, 5, pp. 819-856

Poprawe M. (2015), *On the relationship between corruption and migration: Empirical evidence from a gravity model of migration*, «Public Choice», 163, 3-4, pp. 337-354

Shen J.F. (2015), *Explaining interregional migration changes in China, 1985-2000. Using a decomposition approach*, «Regional Studies», 49, 7, pp. 1176-1192

Shi Y., Zheng X.Q., Sun Y.B. *et al.* (2014), *Intra-urban migration patterns in the perspective of improved radiation model-a case study of Beijing*, «Sustainable Development of Industry and Economy, Pts 1 and 2», 869-870, pp. 110-116

Song W. and Liu M.L. (2014), *Assessment of decoupling between rural settlement area and rural population in China*, «Land Use Policy», 39, pp. 331-341

Song X. and Wang Y. (2005), *Zhongguo renkou kongjian fenbu bianhua de shoulianxing fenxi*, «Zhongguo renkou kexue», 4, pp. 54-58

Stillwell J. (2005), *Inter-regional migration modelling-a review and assessment*, https://core.ac.uk/download/files/153/7046065.pdf

Stillwell J. and Congdon P. (1991), *Migration Models: Macro and Micro Approaches*, Belhaven Press, London

Sun W.K., Wang X.H. and Bai C.E. (2014), *Income inequality and mobility of rural households in China from 2003 to 2006*, «China Agricultural Economic Review», 6, 1, pp. 73-91

Wang G. (1993), *Woguo shengji renkou qianyi yu juli guanxi zhi tantao*, «Population and Economy», 2, pp. 3-8

Wang Y., Chen Y.N. and Li Z. (2013), *Evolvement characteristics of population and economic gravity centers in Tarim River Basin, Uygur autonomous region of Xinjiang, China*, «Chinese Geographical Science», 23, 6, pp. 765-772

Wu Z.M. and Yao S.J. (2003), *Intermigration and intramigration in China: A theoretical and empirical analysis*, «China Economic Review», 14, 4, pp. 371-385

Yang C.F., Han H.R. and Song J.P. (2014), *Spatial distribution of migration and economic development: A case study of Sichuan Province, China*, «Sustainability», 6, 10, pp. 6509-6528

Appendix

TABLE 6 • INTERPROVINCIAL MIGRATION MODEL COEFFICIENTS – WITHOUT TAKING LOGS

	2000		2010	
	B	S.E.	B	S.E.
Origin urban population	-1153.473	1871.119	-2990.402	2612.489
Origin rural population	1620.645***	458.608	3439.309***	841.18
Destination urban population	8543.815***	1871.119	17445.65	17895.62
Destination rural population	-351.475	458.608	857.848	705.603
Distance	-29.193**	8.972	-58.578***	13.768
Origin urban income	-7158.314	7806.751	-24164.92**	7371.542
Origin rural income	5906.455	18647.03	35979.49*	14141.54
Destination urban income	51622.01***	7806.751	38747.6***	7253.301
Destination rural income	-66902.26***	18647.03	-33959.62**	13713.5
Constant	-189019.1***	45326.22	-355741.6**	175037

I tre lavori qui pubblicati sono i vincitori della quarta edizione (2016) del Giorgio Rota Best Paper Award for Young Researchers *sul tema "The Economics of Migration" e sono stati presentati alla* Giorgio Rota Conference *di giugno 2016 tenutasi presso il Campus Einaudi in collaborazione con il Dipartimento di Economia e Statistica 'Cognetti de Martiis' e FIERI.*
Il Premio, istituito dal Centro Einaudi in memoria dell'economista torinese che ne era stato uno degli animatori, è sostenuto dalla Fondazione CRT.

Il Centro di Ricerca e Documentazione Luigi Einaudi conduce attività di ricerca in proprio o su commissione di enti nazionali o internazionali, cura siti web e la pubblicazione di periodici, svolge opera di formazione di giovani studiosi e ricercatori, organizza seminari, conferenze, convegni. L'attività è finanziata attraverso il contributo di soci e sostenitori, nonché con i proventi delle ricerche. Il Centro, che non ha fini di lucro e dal 2006 gode del riconoscimento della Regione Piemonte, è nato a Torino nel 1963 come libera associazione di imprenditori e intellettuali, grazie all'iniziativa e all'impegno di Fulvio Guerrini.

ISBN 978-88-941152-3-9