DISCUSSION PAPER SERIES

IZA

IZA DP No. 10213

Networks and Migrants' Intended Destination

Simone Bertoli
Ilse Ruyssen

September 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Networks and Migrants' Intended Destination

Simone Bertoli
CERDI, University of Auvergne and IZA

Ilse Ruyssen
SHERPPA, Ghent University and IRES, Université Catholique de Louvain

Discussion Paper No. 10213
September 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 10213
September 2016

ABSTRACT

Networks and Migrants' Intended Destination*

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.

JEL Classification: F22

Keywords: international migration, networks, intentions

Corresponding author:

Simone Bertoli
CERDI
University of Auvergne
Bd. F. Mitterrand, 65
F-63000 Clermont-Ferrand
France
E-mail: simone.bertoli@udamail.fr

* 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; 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.

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 (see, for instance, Pedersen *et al.*, 2008; Beine *et al.*, 2011, 2015; Beine and Salomone, 2013; Bertoli and Fernández-Huertas Moraga, 2015). The implicit assumption behind this approach, which reflects binding data constraints, is that all potential migrants equally benefit from the networks at destination.[1] This assumption is at odds with theoretical representations of social networks (see 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 (see 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.[2] 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

[1] 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 identification of network effects from gravity equations on aggregate data on bilateral migration flows.

[2] 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.

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.[3]

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.

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., *un*observed factors that influence both the geographical distribution of

[3] 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.

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.[4] 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.[5] 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.

2.1 Intending migrants

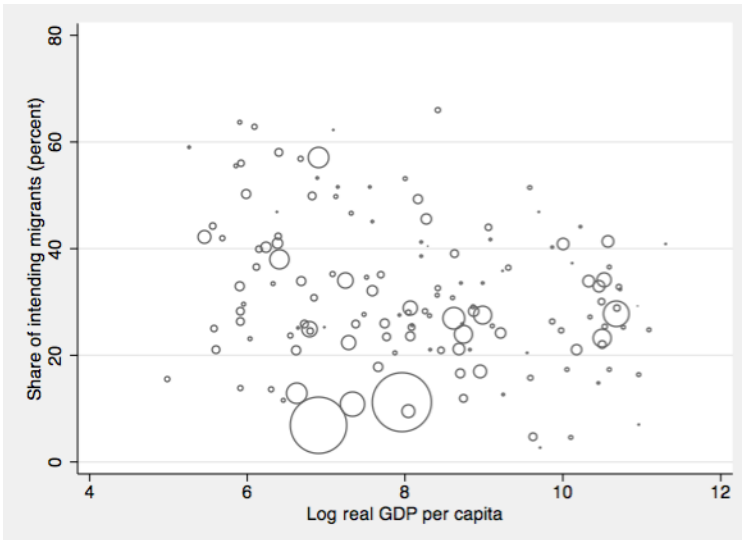
The Gallup World Polls include two related questions on the intention to migrate, asked in all countries between 2007 and 2011: (*i*) "Ideally, if you had the opportunity, would you like to move to another country, or would you prefer to continue living in this country?", and

[4] 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.

[5] 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).

(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.[6]

Figure 1: Share of intending migrants and income per capita



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.[7] The ten countries with the highest shares of intending 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

[6]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.

[7]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.

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).[8] 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.

Table 1 reports the distribution of intending migrants across the top-20 countries of destination.[9] The natives aged 15 to 49 in our sample intend to migrate towards 185 different

[8]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.

[9]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).

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.[10] 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.[11] 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.[12] Thus, questions (*iii*) and (*iv*) give us information about up to three

[10] The first five 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).

[11] 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.

[12] 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 five years?", with information on the

countries in which each individual is directly connected to someone who could provide help to him or her.[13] 58 percent 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 (2015a). 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,[14] and its (weighted) average stands at 0.519, with a standard deviation of 0.099.[15] 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.

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.

[13]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.

[14]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 confindence level are Bahrain (p-value 0.096) and Namibia (0.025).

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

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:

$$U_{ijk} = V_{ijk} + \epsilon_{ijk}, \tag{1}$$

where $V_{ijk} \equiv \mathbf{x}_{ijk}'\boldsymbol{\beta}_{jk}$ 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 $\boldsymbol{\beta}_{jk}$. 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}.

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.[16]

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

[16] "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, p. 85).

in the choice set D_j.[17] 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.[18] 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 $\widehat{\beta}_{1j}^{R_j^n} = \widehat{\beta}_{1j}$.[19]

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.[20] 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 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).

[17]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).

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

[19]See Section 5.2 for more details.

[20]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.

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
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

(*continued*)

Table 2: Sample size and number of intended destinations (*continued*)

Country	Waves	Obs.	Dest.	Country	Waves	Obs.	Dest.
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.

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.

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 $\mathbf{z}_{ij}$ of individual characteristics, including sex, four age cohorts,[21] and a dummy that takes the value one for individuals who completed at least nine years of education.[22] 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).[23] 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.[24]

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 $1 +$

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

[22]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.

[23]We also present specifications where time-varying dyadic dummies, i.e., $\mathbf{d}_{jkt}$, thus controlling also for variations over time in the size of the diaspora.

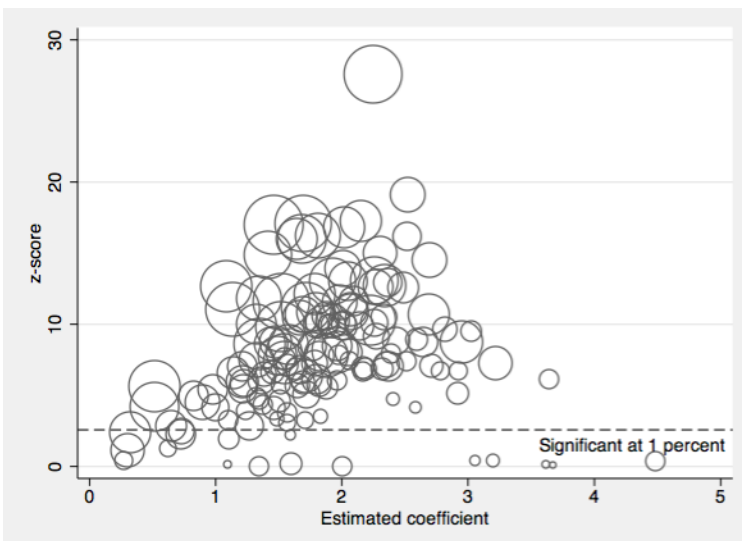
[24]Our specification is actually more general, as it does not require the diaspora to be defined on the basis of the country of birth; for instance, our specification can allow for the attractiveness of the United States for potential Ecuadorian migrants to depend on the size of the diaspora of all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

$6(N_j - 1)$ 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}.[25] 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.

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.

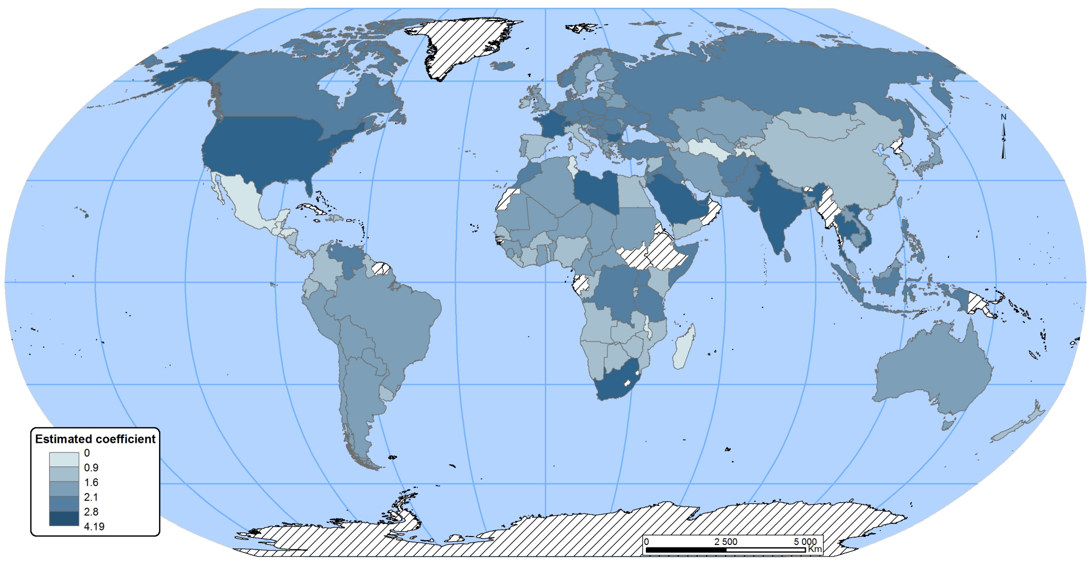
Source: Authors' elaboration on Gallup World Polls and World Bank (2015a,b).

Figure 3 plots the values of the estimated coefficients in a world map, and it reveals that

[25]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.

there is no clear geographical pattern in the values of the estimates for the coefficient of distance-one connections.[26]

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.

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.[27]

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

[26]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.

[27]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.

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 $\mathbf{d}_{jk}$, which reflect the differences in the deterministic component of location-specific utility,[28] 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).[29] 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.

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

[28]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 $\mathbf{z}_{ij}$.

[29]U_{ijk} 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 $E(U_{ijk}|U_{ijk} > U_{ijh}\ , \forall h \in D/\{k\}) = E(U_{ijl}|U_{ijl} > U_{ijh}\ , \forall h \in D/\{k\})$.

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 $\widehat{\beta}_{1j}$ is 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.[30] 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.[31]

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

[30]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.

[31]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.

coefficient for d_{ijk} that lies in the confidence interval of the one obtained from our 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.[32]

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:[33] 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,[34] and an asset index *à la* Dustmann and Okatenko

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

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

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

(2014).[35] 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 distribution of one's own distance-one connections.[36] 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).[37] 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.[38] The *un*observed proficiency in English, which is potentially correlated with the likelihood of having a distance-one connec-

[35] 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.

[36] 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.

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

[38] 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.

tion 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,[39] 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 these countries from the choice set of j-born intending migrants.[40] 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.[41]

This restriction in the choice set does *not* result in a systematic reduction in the estimated effect of distance-one connections, as the ~~(weighted)~~ correlation of the point estimates with those from our benchmark specification stands at 0.391.[42]

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

[40] 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.

[41] 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.

[42] 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

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.

References

Beine, M., F. Docquier, and Ç. Özden (2011): "Diasporas," *Journal of Development Economics*, 95(1), 30–41.

——— (2015): "Dissecting network externalities in international migration," *Journal of Demographic Economics*, 81(4), 379–408.

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

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

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

excluded when computing the correlation.

——— (2015): "The size of the cliff at the border," *Regional Science and Urban Economics*, 51, 1–6.

Bertoli, S., J. Fernández-Huertas Moraga, and F. Ortega (2013): "Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions," *Journal of Development Economics*, 101(1), 75–91.

Carrington, W. J., E. Detragiache , and T. Vishwanath (1996): "Migration with endogenous moving costs," *American Economic Review*, 86(4), 909–930.

Chort, I. (2014): "Mexican Migrants to the US: What Do Unrealized Migration Intentions Tell Us About Gender Inequalities?," *World Development*, 59, 535–552.

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

Comola, M., and M. Mendola (2015): "Formation of Migrant Networks," *The Scandinavian Journal of Economics*, 117(2), 592–618.

Creighton, M. (2013): "The Role of Aspirations in Domestic and International Migration," *Social Science Journal*, 50(1), 79–88.

de Palma, A., and K. Kilani (2007): "Invariance of conditional maximum utility," *Journal of Economic Theory*, 132(1), 137–146.

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

Docquier, F., J. Machado, and K. Sekkat (2015): "Efficiency Gains from Liberalizing Labor Mobility," *The Scandinavian Journal of Economics*, 117(2), 303–346.

Docquier, F., G. Peri, and I. Ruyssen (2014): "The Cross-country Determinants of Potential and Actual Migration," *International Migration Review*, 48, S37–S99.

Dustmann, C., and A. Okatenko (2014): "Out-migration, wealth constraints, and the quality of local amenities," *Journal of Development Economics*, 110, 52–63.

Esipova, N., J. Ray, and A. Pugliese (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., J. Wahba, and Y. Zenou (2014): "Strong Versus Weak Ties in Migration," Discussion Paper No. 10248, London: CEPR.

Grogger, J., and G. H. Hanson (2011): "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 95(1), 42–57.

Guimaraes, P., O. Figueiredo, and D. Woodward (2003): "A Tractable Approach to the Firm Location Decision Problem," *Review of Economics and Statistics*, 85(1), 201–204.

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

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

Manchin, M., R. Manchin, and S. Orazbayev (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, Vienna.

Mayer, T., and S. Zignago (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 *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. New York: Academic Press.

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

Munshi, K. (2003): "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, 118(2), 549–599.

——— (2016): "Community Networks and Migration," in *Oxford Handbook on the Economics of Networks*, ed. by J. Bramoullé, A. Galeotti, and B. W. Rogers, pp. 630–648. Oxford: Oxford University Press.

Munshi, K. and M. Rosenzweig (2016): "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 106(1), 46–98.

Ortega, F., and G. Peri (2013): "The Role of Income and Immigration Policies in Attracting International Migrants," *Migration Studies*, 1(1), 27–46.

Özden, Ç., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011): "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000," *World Bank Economic Review*, 25(1), 12–56.

Patel, K., and F. Vella (2013): "Immigrant networks and their implications for occupational choice and wages," *Review of Economics and Statistics*, 95(4), 1249–1277.

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

Sjaastad, L. A. (1962): "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70(1), 80–93.

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

——— (2015b): *World Development Indicators*. Washington DC: World Bank.

A Appendix

A.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:

$$\text{flow}_{jkt} = \exp\left[\alpha \ln(\text{intention}_{jkt}) + \boldsymbol{\beta}'\mathbf{x}_{jk} + \mathbf{d}_{jt} + \mathbf{d}_{kt} + \epsilon_{jkt}\right] \tag{A.1}$$

where $\mathbf{x}_{jk}$ 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,[43] 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 A.1 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.

[43] 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	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	flow_{jk}	flow_{jk}	flow_{jkt}	flow_{jkt}	flow_{jkt}	flow_{jkt}
$\ln(\text{intentions}_{jkt})$	0.800***	0.627***	0.540***	0.409***	0.444***	0.345***
	[0.048]	[0.038]	[0.028]	[0.027]	[0.032]	[0.033]
$\ln(\text{networks}_{jk})$		0.247***		0.242***		0.192***
		[0.038]		[0.022]		[0.028]
$\ln(\text{distance}_{jk})$	-0.588***	-0.401***	-0.712***	-0.496***	-1.031***	-0.816***
	[0.066]	[0.060]	[0.045]	[0.049]	[0.056]	[0.055]
Contiguity_{jk}	0.585***	0.372**	0.506***	0.314***	1.556***	1.081***
	[0.167]	[0.148]	[0.095]	[0.086]	[0.159]	[0.154]
Common language_{jk}	0.318**	0.371***	0.515***	0.529***	0.583***	0.650***
	[0.130]	[0.119]	[0.073]	[0.068]	[0.087]	[0.091]
Colony_{jk}	0.308**	0.033	0.348***	0.065	0.434***	0.109
	[0.132]	[0.117]	[0.056]	[0.061]	[0.077]	[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; *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level; the dependent variable in specifications (1)-(2) is obtained collapsing the variables for each origin-destination pair over time before taking the logarithmic transformation; specifications (5)-(6) exclude from the sample the origin countries that are classified 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).

A.2 Benchmark estimates

Table A.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

(*continued*)

Table A.2: Estimated coefficients for distance-one connections (*continued*)

Country	obs.	coeff.	s.e.	Country	obs.	coeff.	s.e.
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	$7.1{*}10^4$
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 Territory	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 and 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.