WIFO

ÖSTERREICHISCHES INSTITUT
FÜR WIRTSCHAFTSFORSCHUNG

WORKING PAPERS

Ethnic Networks and the Location Choice of Migrants in Europe

Klaus Nowotny, Dieter Pennerstorfer

415/2011

Ethnic Networks and the Location Choice of Migrants in Europe

Klaus Nowotny, Dieter Pennerstorfer

WIFO Working Papers, No. 415
December 2011

E-mail addresses: Klaus.Nowotny@wifo.ac.at, Dieter.Pennerstorfer@wifo.ac.at
2011/353/W/0

Medieninhaber (Verleger), Hersteller: Österreichisches Institut für Wirtschaftsforschung • 1030 Wien, Arsenal, Objekt 20 • Tel. (43 1) 798 26 01-0 • Fax (43 1) 798 93 86 • http://www.wifo.ac.at/ • Verlags- und Herstellungsort: Wien
Die Working Papers geben nicht notwendigerweise die Meinung des WIFO wieder
Kostenloser Download: http://www.wifo.ac.at/wwa/pubid/43198

Ethnic Networks and the Location Choice of Migrants in Europe

Klaus Nowotny* Dieter Pennerstorfer†

Draft version: December 12, 2011

Abstract

In this paper we analyze the role of ethnic networks in the location decision of migrants to the EU-15 at the regional level. Using a random parameters logit specification we find a substantially positive effect of ethnic networks on the location decision of migrants. The effect is, however, decreasing in network size. Furthermore, we find evidence of spatial spillovers in the effect of ethnic networks: ethnic networks in neighboring regions significantly help to explain migrants' choice of target regions. The positive effects of ethnic networks thus also extend beyond regional and national borders. Analyzing the trade-off between potential income and network size, we find that migrants would require a sizable compensation for living in a region with a smaller ethnic network, especially when considering regions where only few previous migrants from the same country of origin are located.

JEL classification numbers: F22, R23, C35

Keywords: network migration, ethnic networks, random parameters logit

*Corresponding author. Austrian Institute of Economic Research (WIFO), P. O. Box 91, A-1103 Vienna, Austria. E-mail: klaus.nowotny@wifo.ac.at.

†Austrian Institute of Economic Research (WIFO).

1 Introduction

Previous research has shown that migrants' choice of location within a country can be explained by differences in economic opportunities across regions (like higher wages or a higher probability of finding employment, see Davies et al., 2001) or by some regions or cities being "natural hubs" for migrants (they act as "ports of entry" into a country because of infrastructure endowments like sea- or airports or administrative institutions like central immigration offices). But these factors cannot fully explain the observation that migrants tend to settle where other migrants of the same ethnicity migrated before, resulting in a geographic concentration of migrants with similar ethnicity in specific locations. Since a seminal study on ethnic migrant concentration in the U. S. by Bartel (1989), several studies have formulated hypotheses explaining migrant concentrations theoretically (see Massy et al., 1993, for an overview of some earlier work, or Carrington et al., 1996; Gross and Schmitt, 2003; and Chiswick and Miller, 2005) and have identified the importance of ethnic networks for the location decision of migrants (see Zavodny, 1999; Bauer et al., 2000; Gross and Schmitt, 2003; Åslund, 2005; Pedersen et al., 2008; Damm, 2009a). Other studies have highlighted the role of ethnic networks for employment and earnings opportunities or educational attainment (see Cutler and Glaeser, 1997; Munshi, 2003; Cardak and McDonald, 2004; Chiswick and Miller, 2005; Damm, 2009b, to name just a few) or the role of "ethnic capital" in determining economic performance (see Borjas, 1992, 1995).

However, most of the previous literature has focused only on local networks and has not considered the spatial structure of networks in and around a region. But the positive effect of a network may not be limited to regional borders: newly arriving migrants can also benefit from networks in neighboring regions by gaining information on labor market opportunities in these neighboring regions, or by the provision of ethnic goods produced in other regions. Furthermore, some ethnic goods might be provided only for migrants in a certain region if the network size in adjacent regions and in other regions of the country is large enough. Larger networks can thus also be associated with a higher variety of ethnic goods.

We thus contribute to the existing literature by considering not only the size of the local ethnic network as a determinant of migrants' location choice, but also the size of the ethnic network in neighboring regions and other regions of the host country. This is, to the authors' best knowledge, the first article to date which explicitly incorporates this form of spatial heterogeneity in the effect of ethnic networks on migrant's location choice.

Second, we contribute to the existing literature on ethnic networks and the location choice of migrants by analyzing location choice at the regional level for European countries, while other studies deal with this topic either at the national level (Pedersen et al., 2008; Geis et al., 2008) or focus on the regions of a single country (Bartel, 1989; Åslund, 2005; Damm, 2009b). Third, we examine the trade-off between income and ethnic network size and provide a first approximation of the compensating variation for changes in the ethnic network size. Fourth, we derive an empirical estimate of the optimal ethnic network size for migrants in the EU.

As in previous empirical studies (e. g., Davies et al., 2001; Christiadi and Cushing, 2008) the location decisions are estimated at the individual level using a discrete choice model based on random utilities. If our hypothesis is true and networks in neighboring regions matter for the location decision, the independence from irrelevant alternatives property is violated, and the commonly used conditional logit model (McFadden, 1974) is no longer applicable. We therefore follow Gottlieb and Joseph (2006) and apply the more suitable random parameters (mixed) logit framework (see McFadden and Train, 2000).

Based on 2007 data from the European Labour Force Survey our empirical analysis shows that the probability of moving to a region depends not only on the local ethnic network but also—albeit to a smaller extent—on the ethnic network in adjacent regions and other regions of the same country. Ignoring the effects of networks in neighboring regions thus overestimates the effect of network size in the host region and leads to biased results. Besides this, we find the expected effects of economic attributes like region size or regional income and unemployment on the location decision of migrants. Deriving the trade-off between income and ethnic network size, we are able to calculate the Euro value of a variation in ethnic networks. Our results show that ethnic networks are highly important for the location decision, and that migrants would require a sizable compensation for moving to a region with a smaller ethnic network.

2 Literature review

One of the most frequently cited theories explaining ethnic migrant clusters is that migrant networks produce positive externalities for members of the same ethnic group, so that the costs of migration decrease with the number of previous migrants: networks can provide help with the settlement process, decrease the perceived alienation in the host country (Bauer et al., 2000) or provide financial assistance (Munshi, 2003). Furthermore, networks can provide their members with ethnic goods like food, clothing, social organizations, religious services, media (radio, newspapers, etc.) or marriage markets (Chiswick and Miller, 2005), and the provision of these ethnic goods can be expected to increase with the stock of migrants with similar ethnic background. This creates an externality which provides incentives for other immigrants to settle in regions where they can enjoy a larger supply of ethnic goods. If there are economies of scale in the production of ethnic goods (as can, for example, be expected for religious services or media), geographic concentration facilitates the supply of these goods at lower prices and reduces the costs of living (especially if ethnic goods make up a large part of the consumption basket), which attracts more immigrants to move into this region even if they could earn a higher wage somewhere else (Chiswick and Miller, 2005).

Ethnic networks also provide information externalities: by being in contact with previous migrants, new arrivals can benefit from a better availability of information on employment opportunities which increases their labor market prospects (Gross and Schmitt, 2003). New arrivals can also benefit from job referrals by more established members of the network (Munshi, 2003). Furthermore, if employers with migration background prefer to employ other migrants of similar ethnic origin instead of natives (Andersson and Wadensjö, 2007), a separate migrant labor market can emerge which can even sustain a higher wage than the larger "general" labor market (Gross and Schmitt, 2003).[1]

A variety of empirical studies in the literature support the network migration hypothesis and find positive effects of ethnic networks on the location decision of newly arrived migrants. However, most of the previous work focuses on the U. S., while there are only few studies covering European countries. Two notable exceptions are Pedersen et al. (2008), who estimate the determinants for migration flows to 22 OECD countries and find a robust and sizable effect of ethnic networks on migration flows, and Geis et al. (2008), who found networks to have a positive (but decreasing in network size) effect on migrant's choice between four OECD countries (France, Germany, United Kingdom, and the U. S.). Other studies on European countries take a single-country perspective: focusing on Denmark, Damm (2009a) showed that the relocation hazard of refugees randomly assigned to a municipality during the Danish spatial dispersal policy is lower for those assigned to a municipality with a higher percentage of co-nationals, while Åslund (2005) found similar effects for immigrants to Sweden

[1]Edin et al. (2001) found empirical support for a positive effects of ethnic networks on migrant earnings. In an analysis of Mexican migrants in the U. S., Munshi (2003) provides evidence that networks not only increase the probability of employment, but also help to channel network members into higher paying occupations. (Bartel, 1989, p. 388), on the other hand, showed that clustering negatively influences the economic success of migrants. One explanation for this is that migrant clusters are negatively correlated with foreign language fluency (Lazear, 1999), which is in turn a prerequisite for entering the host country's labor market (see also Bauer et al., 2005). Damm (2009b) concludes that the positive effects of ethnic networks more than outweigh the negative effects, and that all things considered living in a region with a larger ethnic network has a positive effect on wages.

subject to the "Whole of Sweden Strategy" as well as a preference of migrants for regions with larger ethnic networks before the implementation of the strategy.

While there is strong evidence that ethnic migrant networks have a positive effect on the location decision, there can also be negative effects on the utility of both previous migrants (Heitmueller, 2006) and prospective new arrivals: continuing migration reduces the income differentials between sending and receiving countries and the wages of migrant cohorts. A similar effect will arise if housing prices increase following an influx of migrants into a region. This negative effect of decreasing wages and/or increasing housing prices will at some point dominate the positive network externality effect, leading to a decline in the attractiveness of a formerly popular ethnic cluster (Portnov, 1999). There will thus be an optimal size of the regional network beyond which every new migrant decreases the utility of previous migrants already living in the region. If prospective emigrants take this into consideration when deciding where to locate, an inversely U-shaped effect of network size on the probability of moving to a specific region can arise (Bauer et al., 2002).[2]

As an alternative to network effects, Epstein (2002) and Bauer et al. (2005) argued that herd behavior can constitute another explanation for the creation of ethnic clusters in specific regions. Herd migration occurs if there is imperfect information as to which among alternative target locations provides the highest utility. If a potential migrant observes only the outcome of previous migrants' destination choices, but not the "signal" that determined their choice, she might discount her private information about alternative target regions and follow the flow of previous migrants in the belief that they must have had information not available to her.[3]

3 Data and econometric framework

In the empirical analysis we will test for the importance of ethnic networks in the location decision of migrants in Europe. The data on migrants and ethnic networks used in the empirical analysis is taken from a special evaluation

[2]Local ethnic networks can, however, still grow beyond this optimal size if the region still provides a higher utility compared to all other available regions, even if new migrants take into account that their utility will decrease with every other migrant that follows (Bauer et al., 2002; Heitmueller, 2006).

[3]Herd behavior can lead to inefficiencies if previous migrants also discounted their private information in the belief that those who went there before them had information they do not have, while they could have gained a higher utility by following their private information (which must, however, not be the location with the objectively best conditions either). Herd behavior and network effects are—although conceptually different—not mutually exclusive: both effects can exist simultaneously and determine the location decisions of migrants. The presence of network externalities in this context can even increase the probability that herd behavior will be observed (Epstein, 2002).

of the 2007 European Labour Force Survey (EU-LFS). The EU-LFS is a regular questionnaire surveyed among a representative sample of households in all countries of the EU-27. The special evaluation of the 2007 EU-LFS available to us provides not only information on the region of residence (at the NUTS-2 level), but also detailed information on the nationality and the country of birth of individuals living in the EU. We use country of birth to define ethnicity (so all individuals born in the same source country are assumed to belong to the same ethnic group) and consider all individuals born outside their country of residence to be "migrants".

The data allow us to differentiate between those who moved during the last 10 years (between 1998 and 2007) and those who have been living in the same country for more than 10 years. The location choice is only modeled for those who migrated between 1998 and 2007 still living in their host country at the time of the interview (there is no information on repeat and return migration). We focus on the 15 EU member states as of 1998, but due to missing information on the country of birth for Germany and Ireland only 13 can be considered as host countries.[4] Although the EU-LFS is a survey, the detailed survey information is not available to us, so observations are weighted according to the weight attached to them in the LFS. The data is thus used like administrative data in grouped regressions. All in all, we model the location decision of 8,988,710 migrants age 15 or older from 166 different countries who migrated to the 13 EU countries considered during the 1998–2007 period. We assume that a migrant k's choice set R consists of 158 NUTS-2 regions in these 13 countries[5] and that this choice set is exhaustive.[6]

3.1 A random utility approach to location choice

To model the location decision of migrant k, a random utility framework can be applied (Marschak, 1960) where each region $r \in R$ yields a region-specific utility U_{kr}. We impose the simple behavioral model of a utility-maximizing decision

[4] Austria, Belgium, Denmark, Finland, France, Greece, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom.

[5] Overseas territories as well as the Spanish exclaves Ceuta and Melilla are not considered. The same holds true for the relatively remote Canary Islands and the Azores and Madeira island regions. Åland (Finland) as well as the Highlands and Islands and North Eastern Scotland regions in the U. K. must be excluded because of lacking data. Because of data restrictions, Denmark is treated as a single NUTS-2 region and Serbia, Montenegro and the Kosovo are considered a single source country.

[6] The choice of migrating vs. staying in the home country is not modeled because this would imply both modeling the choice of all "stayers" in all source countries as well as modeling the choice of all migrants from all source countries to all other countries. Since this is practically infeasible, it is assumed that migrants have already decided to migrate to the 13 EU countries under investigation and then choose a location from among the 158 regions in these countries. Migrants moving between the 13 host countries are also excluded, because for them the regions of their home country would be included in the choice set R while they would actually not be allowed to choose one of these regions because they would then not be regarded as migrants.

maker: migrant k chooses alternative $s \in R$ if and only if $U_{ks} > U_{kr} \; \forall \; s \neq r$. Because the decision maker's utility is not known, observable characteristics of the alternatives X_{kr} can be used to define the representative utility $V_{kr} = V(X_{kr}) \; \forall \; r$ which is a function of alternative-specific variables like measures for ethnic networks in r and in adjacent regions. Assuming that representative utility is linear in the attributes of the alternatives, the utility function is given by

$$U_{kr} = V_{kr} + \varepsilon_{kr} = \beta' X_{kr} + \varepsilon_{kr} \tag{1}$$

ε_{kr} is unknown and treated as random. The final outcome can thus only be predicted in terms of probability.

3.2 Ethnic networks

Our main variable of interest to be included in X_{kr} is the size of the ethnic network. For a migrant in ethnic group j, ethnic network size in a specific region s (the local ethnic network in region s) is defined as

$$\text{Network}_{jr} = \frac{m_{js}^{10+}}{M_j^{10+}}$$

where $M_j^{10+} = \sum_{r=1}^{R} m_{jr}^{10+}$ is the sum of migrants of ethnic group j who have been living in all regions for more than 10 years. Our network variable thus measures the proportion of migrants of the same ethnicity who have been living in region s for at least 10 years relative to all migrants of the same ethnicity who have been living in the 13 EU countries considered for at least 10 years. This definition assumes that migrants are mostly interested in the region which hosts the largest ethnic network, irrespective of its absolute size.[7] As the summary statistics of table 1 show, the average network size is 6.7 %, but network size can vary from zero to 100 percent. Because the marginal utility of networks can decrease with network size the squared network variable is also included in the regression.

As outlined in the introduction, the positive effect of ethnic networks does not necessarily end at the region's border. E. g., ethnic goods can also be con-

[7] This definition is chosen because the absolute network size does not take the total population of an ethnic group into consideration: if a region hosts a network of 1.000 migrants of ethnic group A and 10.000 migrants of ethnic group B, the region would be considered more attractive for members of B than for members of A. But if the total number of migrants from ethnic group A in all regions is only 2.000 while the total population of ethnic group B in all regions is 200.000, the region would clearly be of higher importance to members of group A than members of group B. The relative network size takes this into account and places a higher weight on the region for members of group A (network size: 50 %) than for members of group B (network size: 5 %). Section 5.1 relaxes this assumption and estimates the model using absolute network size.

sumed by individuals living in neighboring regions, or migrants could live in one region and commute to work in a neighboring region where the ethnic network will help them find employment. If there are spatial spillovers, the individual may not only be concerned with local network size in his intended target region, but also with the size of the ethnic network in neighboring regions. We therefore also include the sum of the ethnic networks in neighboring regions as an additional variable in the regression:

$$\text{Network}_{js}^{N_1} = \frac{\sum_{l_1^s=1}^{L_1^s} m_{jl_1^s}^{10+}}{M_j^{10+}}$$

with $L_1^s \subset R$ the set of regions sharing a border with region s. Furthermore, we also include the sum of the networks in second neighbor regions L_2^s (the neighbors of the L_1^s regions, except s and regions summarized in L_1^s):

$$\text{Network}_{js}^{N_2} = \frac{\sum_{l_2^s=1}^{L_2^s} m_{jl_2^s}^{10+}}{M_j^{10+}}$$

To the authors' best knowledge this is the first article to date which explicitly incorporates this form of spatial heterogeneity in the context of migrant's location choice.[8]

If there are ethnic goods with strong economies of scale in production which must be provided at the national level to be produced efficiently, the total network in the host country also affects the location decision. We therefore also include the sum of the ethnic networks in the rest of the host country (L_C^s):

$$\text{Network}_{js}^{N_C} = \frac{\sum_{l_C^s=1}^{L_C^s} m_{jl_C^s}^{10+}}{M_j^{10+}}$$

We thus consider network effects both at the regional as well as the national level. For networks in neighboring regions we differentiate between neigboring regions in the same country and adjacent regions in neighboring countries. While the set L_1^s contains only neighboring regions within the same country, adjacent regions in neighboring countries are included in an alternative set ${L'}_1^s$ with the size of the ethnic network in regions ${L'}_1^s$, $\text{Network}_{js}^{N_1'}$, defined as above. The same applies to ethnic networks in second neighbor regions of neighboring countries, $\text{Network}_{js}^{N_2'}$.

[8]We focus on the spatially lagged network of migrants who moved into the region more than 10 years ago (which is exogenous in the regression), and not so much on spatially contemporaneous dependence (i. e., spatial lags or spatial errors, see Anselin, 2006), as it will take some time for a new arrival to provide things like ethnic goods or information externalities to other members of the network. Besides, there are (to the best knowledge of the authors) no estimators allowing for contemporaneous spatial dependence in models of this kind.

A priori it can be expected that these variables affect the location choice of migrants at a lesser extent, if at all. Migrants of the same ethnicity living in adjacent regions of a neighboring country will not be able to help with immigration issues and bureaucratic structures because of national differences in migration regimes and procedures. Furthermore, labor and housing markets in different countries are subject to different laws, making positive network externalities rather unlikely. National borders will, however, play a lesser role for the consumption of ethnic goods because there are no restrictions on trade and cross-border mobility among EU countries. If significant, the coefficients can, in comparison with their within-country counterparts, provide information about border effects in network externalities.

[Table 1 about here.]

3.3 Other explanatory variables

Our choice of other explanatory variables included in X_{kr} follows other studies on this topic (see Bartel, 1989; or Davies et al., 2001). Two types of variables will be added to the regression: variables which are specific to the region or country of residence, as well as country-pair specific variables. As will become obvious from the discussions in section 3.4, variables specific to the source countries (such as unemployment or wage levels, or sending country fixed effects) cannot be considered in the regressions because they do not vary over alternatives. The same holds true for individual characteristics like age or gender.[9]

Among the region specific X_{kr} attributes assumed to influence the probability of moving to a region is the area (measured in 1,000 km^2): even if there is a completely uniform distribution of migrants across all regions, larger regions are more likely to attract larger inflows of migrants. A similar argument can be made for the population (in 100,000): after controlling for region size (area), regions with a higher population share should also attract a higher share of migrants. To control for differences in economic opportunities, we include the unemployment rate (in percent) as well as the average annual income per employee (in € 1,000). Data on population and unemployment (in 2006) as well as average annual income (in 2004) are taken from Eurostat. Regional unemployment rates ranged from 2.3 to 20.2 % in 2006 with an average unemployment rate of 7.3 % (see table 1). The average annual income per employed person was €27,300, and ranged from €10,600 ("Centro" region, Portugal) to €100,000 (Inner London, UK). We also include a dummy variable for regions

[9]These variables could only be included by interacting them with all other variables in the model. However, because of technical and practical limitations on the number of random parameters which can be estimated in a random parameters logit model (see section 3.4), the scope for including individual variables is rather limited and will be left to future research.

which comprise national capitals. Capitals can be expected to receive a ceteris paribus higher share of migrants because they are the cultural, political and administrative centers of the respective countries. We expect a negative effect of the unemployment rate and a positive effect of average annual income on the probability of choosing a specific region. To control for national differences in laws regarding immigration, labor market access as well as other country-fixed effects, dummies for the receiving countries are included (see also Davies et al., 2001).[10]

Among the country-pair specific X_{kr} attributes we include a dummy variable for linguistic closeness taken from CEPII which measures whether a migrant's home and host country share an official language (1, zero otherwise). A common language not only reduces the costs of migration considerably (see Pedersen et al., 2008), but it can also raise the returns-to-skill in the host country (Grogger and Hanson, 2008). We also include a neighborship dummy which is 1 if the host and home countries share a common border, and zero otherwise. Again, a positive effect can be expected because a common border facilitates not only legal, but also illegal immigration and can thus lead to ceteris paribus higher migration. (Former) colonial ties between two countries can also affect the location choice of migrants. Data on colonial relationships are also available from CEPII and we include a dummy variable capturing whether two countries were in a colonial relationship after 1945 (= 1, zero otherwise). To proxy for the costs of migration (or the costs of visiting relatives at home), the distance (in 1,000 km, as the crow flies) between the capital of the migrants' country of origin and the geographical center of the region she lives in and squared distance are also included. For distance, a negative (but possibly decreasing) effect can be expected.

Representative utility V_{kr} is thus assumed to be a linear function of host region specific variables (ethnic networks, area, population, average income, unemployment, capital city dummy variable), host country specific variables (country dummies) as well as country-pair specific variables (common official language, common border, colonial ties after 1945, distance). These determinants can be used as explanatory variables to estimate the location choice of migrants. Descriptive statistics for the independent variables are summarized in table 1.

[10]Although including alternative specific dummy variables is more common in applications of this kind, we include only country specific dummy variables, mainly because of practical reasons: with 158 alternatives, we would have to consider 157 dummy variables, which would not only increase the number of parameters to be estimated considerably, but can also lead to problems with achieving convergence when estimating the model. Furthermore, because in the European Union laws regarding immigration and labor market access—which can be considered decisive for immigrants—do not vary within countries, we believe that country dummies are sufficient to estimate alternative specific fixed effects.

3.4 Econometric method

Assuming that the random utility term ε_{kr} in equation (1) is i.i.d. extreme value, the probability that individual k chooses location s could be estimated by a conditional logit (CL) model (McFadden, 1974).[11] As is well known, in the conditional logit model the odds ratio between two alternatives s and t depends only on the characteristics of s and t and not on the availability or characteristics of other alternatives, a property known as "independence from irrelevant alternatives" (IIA):

$$\frac{P_{ks}}{P_{kt}} = \frac{\exp\left(\beta' X_{ks}\right) / \sum_{r=1}^{R} \exp\left(\beta' X_{kr}\right)}{\exp\left(\beta' X_{kt}\right) / \sum_{r=1}^{R} \exp\left(\beta' X_{kr}\right)} = \frac{\exp\left(\beta' X_{ks}\right)}{\exp\left(\beta' X_{kt}\right)} \tag{2}$$

While IIA has some advantages if satisfied, it will be violated if there are spatial spillovers of ethnic networks: if this is the case, the probability of choosing a specific region s not only depends on the characteristics of s, but also on the characteristics of a set of neighboring regions $R(s) = \{L_1^s, L_2^s, L_C^s, {L'}_1^s, {L'}_2^s\} \subset R$. Similarly, the probability of choosing another region t will not only depend on the attributes of this region but also on the attributes of a subset $R(t) = \{L_1^t, L_2^t, L_C^t, {L'}_1^t, {L'}_2^t\} \subset R$ of t's neighbors. The odds between s and t are then given by

$$\frac{P_{ks}}{P_{kt}} = \frac{\exp\left(\beta_1' X_{ks} + \beta_2' X_{kR(s)}\right)}{\exp\left(\beta_1' X_{kt} + \beta_2' X_{kR(t)}\right)}$$

which violates the IIA property: the ratio of the probabilities no longer depends on the characteristics of s and t alone, but also on the characteristics of the regions in $R(s)$ and $R(t)$.[12]

This calls for a model which does not exhibit the IIA property. Probably the most flexible model is the random parameters logit (RPL, also called mixed or random coefficients logit, see McFadden and Train, 2000; Hensher and Greene, 2003; Train, 2009, and the references contained therein for an overview).[13] Although the random parameters logit framework goes back to the early 1980's (among the first applications are Boyd and Mellman, 1980, and Cardell and Dunbar, 1980) and recent advances in simulation techniques (foremost, the use of Halton draws, see below) and computing power have made its estimation more practicable, applications of the random parameters logit model are still

[11] See also Bartel (1989), Bauer et al. (2000, 2002, 2005), Gottlieb and Joseph (2006), Jaeger (2007) or Christiadi and Cushing (2008) for related applications of the conditional logit.

[12] Including the network size of neighboring regions is thus also a test for IIA in our model of location choice (see also Train, 2009, p. 49).

[13] A probably more common alternative model which relaxes the IIA assumption is the nested logit model. However, while nested logit does not impose IIA between nests, alternatives within a given nest are still assumed to exhibit independence from irrelevant alternatives. The model is thus less flexible than the random parameters logit and therefore not considered.

scarce in migration research (one notable exception is the paper by Gottlieb and Joseph, 2006).

The random parameters model can be derived from utility-maximizing behavior by allowing the parameters of the characteristics X_{kr} in the representative utility function to vary over individuals:[14]

$$U_{kr} = \beta_k' X_{kr} + \varepsilon_{kr}$$

In this utility function, β_k is a vector of coefficients for individual k representing k's preferences. The utility function is thus heterogeneous across individuals, and the coefficient of a regional characteristic can not only have a different magnitudes for different individuals, but also a different sign. The coefficients in β_k are assumed to vary over decision makers with density $f(\beta|\theta)$, where θ are the parameters describing the density of β. As in the conditional logit model, ε_{kr} is assumed to be i.i.d. and follow an extreme value distribution. If the β_k's were known, the probability of choosing a specific region s would be given by:

$$L_{ks}(\beta_k) = \frac{\exp\left(\beta_k' X_{ks}\right)}{\sum_{r=1}^{R} \exp\left(\beta_k' X_{kr}\right)} \tag{3}$$

However, because the β_k's are unobserved and we can therefore not condition on β, the probability of choosing s is the integral of $L_{ks}(\beta_k)$ over all possible values of β_k (Train, 2009, p. 138):

$$P_{ks} = \int \left(\frac{\exp\left(\beta_k' X_{ks}\right)}{\sum_{r=1}^{R} \exp\left(\beta_k' X_{kr}\right)} \right) f(\beta|\theta) d\beta \tag{4}$$

The probability P_{ks} in the random parameters logit is thus the weighted average of the logit formula evaluated at different values of β, with the weights given by the mixing distribution $f(\beta|\theta)$. Because the integral in (4) does not have a closed form solution, it must be approximated through simulation. Simulation is based on drawing a value of β from $f(\beta|\theta)$ and using this draw to calculate the logit probability in (3). This step is repeated many times, and the average computed value of $L_{ks}(\beta_k)$ gives the simulated probability $\check{P}_{ks}$ which can be inserted into the simulated log likelihood

$$SLL(\theta) = \sum_{k=1}^{K} \sum_{s=1}^{R} y_{ks} \ln \check{P}_{ks} \tag{5}$$

[14]An alternative interpretation of the random parameters logit is based on the error components creating correlations among utilities for different alternatives, which is formally equivalent to this interpretation, see Train (2009), p. 139f.

The maximum simulated likelihood estimator is the value of θ that maximizes the simulated log likelihood (see Train, 2009, 144) and can be estimated for example in the STATA statistics package using the estimator by Hole (2007).

The mixing distribution in the random parameters logit $f(\beta|\theta)$ can be normal, lognormal, uniform, etc. If the parameters are assumed to be normally distributed, the estimated θ are the mean and standard deviation of a normal distribution which describe the distribution of a parameter in the population. In our econometric model we follow Gottlieb and Joseph (2006) by specifying some coefficients as fixed and the rest as normally distributed.[15] A fixed parameter is essentially a parameter whose standard deviation is zero (Hensher and Greene, 2003), and for which only a mean will be estimated. We assume the coefficient of area (in 1,000 km^2) to be fixed and the same for all individuals: if migrants were evenly distributed across space, larger regions would have a ceteris paribus higher probability of being chosen, independent of individual tastes. The country-specific dummy variables are also treated as being fixed.[16] All other coefficients are unrestricted and assumed to be normally distributed.[17] The estimated parameters θ for these coefficients are thus the mean and standard deviation of a normal distribution. This also allows us to calculate the area of the density function $f(\beta|\theta)$ which is below and above zero. As mentioned above, in the random parameters logit a coefficient is not necessarily positive or negative for all individuals. If part of the area of $f(\beta|\theta)$ is below zero, a variable constitutes an attractor for some, and a repellent for other individuals.

Our simulation uses quasi-random Halton sequences (Halton, 1960), which is considered more effective than simulation based on random draws (see Bhat, 2001; Train, 1999; Hensher, 2001). Train (2009, p. 230) notes that "[...] a researcher can expect to be closer to the expected values of the estimates using 100 Halton draws than 1000 random draws". Although there is no general agreement on the number of Halton draws to be used to achieve stable parameters, Hensher and Greene (2003, p. 154) note that models with a small number of alternatives and random variables can "produce stability with as low as 25"

[15] Revelt and Train (1998) and Train (1999) cite Ruud (1996) showing that random parameters logit models have a tendency to be unstable when all coefficients are treated as random. Therefore, some coefficients should be fixed.

[16] Although heterogeneity of tastes can be expected as regards to individual's preferences for single countries, the maximum dimension of the Mata routine to generate the Halton draws in the STATA statistics package (see Drukker and Gates, 2006) is 20, so that no more than 20 unrestricted coefficients in β can be estimated.

[17] Although sign restrictions could be imposed by specifying some of the coefficients as being lognormally distributed—for example, the coefficient of income can be expected to be positive for all individuals, although its magnitude may vary between decision makers—we specify the random parameters to be normally distributed to make our model as flexible as possible. Furthermore, lognormal distributions usually have a long right-hand tail, which might be problematic in calculations of the willingness-to-pay or the compensating variation because it often leads to unrealistic mean values (see Hensher and Greene, 2003, for a discussion). The use of the log-normal distribution is also discouraged by Sillano and Ortúzar (2005).

Halton draws per observation, and that "100 appears to be a 'good' number". However, the number of required draws will be higher the more complex the model (Hensher and Greene, 2003, p. 154), so that these results cannot be generalized. We use 500 Halton draws for the simulation of the random parameters logit model.[18]

4 Results

Table 2 shows the results of the random parameters logit regression estimating the location choice of those migrants who migrated to the 13 host countries considered between 1998 and 2007. In addition to the mean and standard deviation of the estimated random parameters which define the normal distribution of the coefficients in the population, table 2 also shows the proportion of this normal distribution which above zero (i. e., the percentage of the population for which the parameter is positive). The fifth column gives the exponentiated coefficients of the random parameters logit, which can be interpreted as mean odds ratios. Finally, the last two columns give the coefficients and odds ratios of a conditional logit regression. Although the conditional logit's IIA assumption is violated if our hypothesis of spatial spillovers in network effects is correct, the conditional logit can still serve as an approximation to a model which relaxes this assumption (cf. Dahlberg and Eklöf, 2003).[19]

[Table 2 about here.]

The results of the random parameters logit support our hypotheses: not only does a larger ethnic network attract more migrants to a region, the estimated probability of choosing a specific region also increases with ethnic networks in neighboring regions. All else equal, at the mean parameter value the odds of choosing a region are 45.6 % larger if the total share of individuals from the same ethnic background in the region increases by 1 percentage point (p. p.). The effect of network size is, however, decreasing, as indicated by the negative effects of the squared network variable. This lends support to the optimal network size hypothesis beyond which the attractiveness of a region declines (see section 4.2). Furthermore, at the mean parameter value the odds ratio is

[18]Halton sequences are usually defined in terms of a prime number. For the simulation of an integral of dimension ι (where the dimension is equal to the number of random parameters), the first ι prime numbers are conventionally used to create ι sequences (Cappellari and Jenkins, 2006). Because the initial elements of the sequences can be highly correlated across dimensions, Train (2009, p. 227) recommends to discard at least the the first κ elements, where κ should be as least as large as the prime number used in the ι'th dimension. Because our model uses 16 random parameters, the first 53 elements are dropped. The model was also estimated using 100, 200, 300 and 400 Halton draws. The parameters tend to stabilize after using 300 draws.

[19]Country fixed effects are not reported in table 2 due to lack of space. The country fixed effects of the random parameters estimations are reported in table A2 in the appendix.

5.3 % larger if the ethnic network in neighboring regions increases by 1 p. p., and even a 1 p. p. increase in the ethnic network of second neighbors is still associated with a change in the relative odds of 3.7 % at the mean parameter. Networks in the rest of the country also play a role for the location decision, but the effect is rather small. Ethnic networks in neighboring regions of other countries also affect the location decision positively, but the estimated coefficients are smaller than for within-country neighbors, which points to substantial border effects in the influence of ethnic networks.

Excluding the spatially lagged network size variables increases the (mean of the) coefficient of the local network variable from 0.376 to 0.651 and reduces the coefficient of the squared network variable from −0.017 to −0.044 in a RPL regression. This indicates that the effect of local ethnic networks is overestimated if spatially lagged network size is ignored, leading to biased results.

The random parameters logit also shows that ethnic networks are an attractor for all individuals: 100 % of the normal distributions of most estimated coefficients are above zero. The only exceptions are the coefficient of the size of the ethnic network in the rest of the country, which is negative for about 23.1 %, as well as the parameter of network size in second neighbors of another country, which is negative for a small proportion of migrants. In addition, for about 10.2 % of migrants the coefficient of the squared network term is positive, indicating that the utility of the individuals increases exponentially with ethnic network size.

Concerning the other variables, the RPL regression shows that migrants ceteris paribus prefer regions with more inhabitants, lower unemployment rates and higher average income. The effect of regional size is negative, but rather negligible in both specifications. As expected, distance—our proxy for the costs of migration—has a negative, but decreasing effect on the location decision. A common official language increases the odds of choosing a specific region, but only for about three quarters (75.4 %) of the migrants. A past colonial relationship between the source and target countries on the other hand affects location choice negatively for most migrants, but about 11.0 % of the migrants to the 13 EU countries considered actually derive a positive utility from living in a region of the former colonizer.

Comparing the results of the random parameters to the conditional logit regression shows that the differences between the mean RPL estimates and the conditional logit are quite substantial for some coefficients, especially those of the network size variable and parameters with a high degree of heterogeneity in the population (such as the capital, common border, and colonial relationship dummies). The evidence provided in this paper does thus not lend support to the hypothesis that a CL model can be used as an approximation to the RPL model which relaxes the IIA assumption, but rather shows that imposing a conditional logit (which implies fixed coefficients) on a empirical model characterized by a high degree of heterogeneity in the coefficients can lead to a severe bias. For example, regions with capital cities exert a positive influence on the location decision of only about 30.0 % of migrants, and the odds ratio of the capital dummy variable is considerably smaller in the RPL than in the CL regression. In another example, a common border has a positive effect on the location decision in the CL model, while the mean RPL estimate is negative: only 46.1 % prefer regions in neighboring countries.

4.1 Compensating variation

The ratio of two parameters in a logit model can be used to calculate the trade-off between two variables x_{1kr} and x_{2kr} (see Davies et al., 2001; Train, 2009). Setting the total derivative of the logit probability to zero and solving for the change in x_{1kr} that keeps the probability of choosing region r constant following a change in x_{2kr} yields:

$$\left.\frac{dx_{1kr}}{dx_{2kr}}\right|_{dP_{kr}=0} = -\frac{\beta_{2k}P_{kr}(1-P_{kr})}{\beta_{1k}P_{kr}(1-P_{kr})} = -\frac{\beta_{2k}}{\beta_{1k}} \tag{6}$$

Using a cost or income measure as x_{1kr} this trade-off can be interpreted as the willingness-to-pay or compensating variation (CV, see Dahlberg and Eklöf, 2003; Sillano and Ortúzar, 2005): if x_{1kr} is income, the ratio gives the amount of money which would compensate an individual for a increase in x_{2kr} by one unit. If both parameters are positive, the compensation for an increase in x_{2kr} is negative, as expected.

The ratio in (6) can be used to calculate the amount of income which would compensate an individual for moving to a region with a smaller ethnic network or, equivalently, the amount of income the individual would be willing to forgo in order to live in a region with a larger ethnic network. Because the parameters in the RPL are random variables which vary across the population with density $f(\beta|\theta)$ we follow Sillano and Ortúzar (2005) and calculate the compensating variation from the individual-level parameters derived from the simulation model (see Train, 2009). Migrant k's compensating variation for a change in the size of the ethnic network can then be calculated as

$$\mathrm{CV}_k(\mathrm{Network}_{js}) = -\frac{\gamma_{1k} + 2\gamma_{2k}\mathrm{Network}_{js}}{\mu_k}$$

where γ_{1k} is individual k's coefficient of the network variable, γ_{2k} her coefficient of the squared network variable and μ_k her coefficient of the average income variable. The compensating variation therefore depends on the size of the network, and will decrease with network size if the squared network parameter is negative (as is the case for about 90 % in our sample, see table 2). Table 3 shows the compensating variation calculated from individual level parameters at different network sizes as well as the calculated compensating variation for ethnic networks in neighboring regions and the rest of the country. They give the amount of annual income an individual would require as compensation for moving to a region where the ethnic network is 1 p. p. smaller.

[Table 3 about here.]

The calculation based on individual level parameters reveals a sizable compensating variation at small network sizes: the amount which would compensate an individual for a 1 p. p. lower network size is, on average, about € 21,600 at a network size of 1 %. Thus, when choosing between two otherwise equivalent regions where one region has a network size of 1 % and the other a network size of 0 %, the probability of moving to these regions would only be equal to the individual if the expected annual income in the region without an ethnic network is € 21,600 higher than the expected income in the other region. This compensating variation is only slightly lower than the mean average annual income per employee in (about € 27,300, see table 1). This implies that regions without networks are highly unattractive, and that ethnic networks are so important that regions without an ethnic network must provide considerably better income opportunities to be considered equally attractive.

As the network size increases, the compensating variation decreases considerably. At the same time, the standard deviation of the CV estimates increases with network size, which reflects the considerable heterogeneity in the individual squared network parameters. At a network size of 5 % the compensating variation drops to about € 13,000 on average, and to about € 2,300 at a network size of 10 %. At the mean network size of 6.65 % the average compensating variation is about € 9,500. The distribution of the compensating variation at the mean network size is depicted in figure 1.

[Figure 1 about here.]

Table 3 also shows that the compensating variation for (the sum of) the network sizes of neighboring regions is considerably smaller than the CV for the network within a region. Furthermore, the compensating variation decreases with distance, and there is a sizable difference between the CV for networks in neighboring regions compared to regions in the rest of the country. In addition, there is also a border effect: the compensating variation for networks in neighboring regions of the same country is about € 3,200, while it is only € 2.300 for neighboring regions in another country. The same pattern holds for networks in second neighbor regions. These results show that the importance of networks decreases with distance to the region of residence, and that there are sizable border effects in the spatial spillovers of ethnic networks.

4.2 Optimal network size

The individual level parameters can also be used to calculate the optimal network size. Differentiating equation (3) with respect to the network variable and solving for the network variable yields the optimal network size:

$$\text{Network}^*_{jr} = -\frac{\gamma_{1k}}{2\gamma_{2k}} \tag{7}$$

Ignoring the 10.2 % for which the squared network variable is positive (and for which the optimal network size would be 100 %, see table 2), the average optimal network size calculated from individual level parameters is about 16.6 %.

Figure 2 shows the distribution of the optimal network size calculated from individual level parameters for those with $\gamma_{2k} < 0$. The distribution of the optimal network size is heavily skewed to the right (see figure 2), and despite excluding individuals with positive parameters for the squared network size variable the distribution of optimal network sizes includes some very large values, some even exceeding 100 % (calculated optimal network size values exceeding 100 % are included in the rightmost category of the histogram). The median optimal network size of 9.6 %, which is about 3.6 times the median actual network size, therefore gives a better representation of the distribution of optimal network sizes than the mean optimal netwzrk size. The smallest optimal network size is 6.2%.

[Figure 2 about here.]

Despite the evidence provided here that there is an optimal ethnic network size beyond which the probability of moving to a region actually decreases with network size, the optimal network size is hardly approached in reality: only 0.8 % of all migrants live in a region where the actual ethnic network is larger than the individual optimal network size, and for only 8.4 % the optimal network size is within ±20 % of the actual network size. The optimal network size is therefore rather hypothetical and hardly exceeded in reality.

5 Robustness

5.1 Alternative definition of network size

For the estimation in section 4, the ethnic network was defined as the percentage of migrants born in the same country of origin who have been living in the same region for 10 years or longer among all migrants from the same ethnic group who have been living in the 13 EU countries considered for at least 10 years. This definition was chosen based on the assumption that migrants focus on the region with the largest ethnic network (relative to the total number of migrants of the same ethnicity, see also footnote 7). This definition of the network variable is debatable. We therefore also consider the absolute size of the ethnic network—the absolute number of migrants of the same ethnicity who have been living in the region for more than 10 years, m_{js}^{10+}—in an alternative specification.The absolute networks in the neighboring regions of the same country are defined accordingly as

$$\text{Absolute network}_{js}^{N_1} = \sum_{l_1^s=1}^{L_1^s} m_{jl_1^s}^{10+}$$

$$\text{Absolute network}_{js}^{N_2} = \sum_{l_2^s=1}^{L_2^s} m_{jl_2^s}^{10+}$$

and the absolute network in the rest of the country as

$$\text{Absolute network}_{js}^{N_C} = \sum_{l_C^s=1}^{L_C^s} m_{jl_C^s}^{10+}$$

Again we differentiate between absolute networks in neighboring regions of the same country and absolute networks in neighboring regions of other countries, and the latter are given by

$$\text{Absolute network}_{js}^{N_1'} = \sum_{l'^s_1=1}^{L'^s_1} m_{jl'^s_1}^{10+}$$

$$\text{Absolute network}_{js}^{N_2'} = \sum_{l'^s_2=1}^{L'^s_2} m_{jl'^s_2}^{10+}$$

Summary statistics are displayed in table 4. On average, a local ethnic network consists of about 14,300 individuals. Algerians in the Île de France region (which includes the French capital, Paris) constitute the largest absolute local ethnic network: 266,000 Algerians had been living in the Île de France for at least 10 years in 2007.

[Table 4 about here.]

We do not consider the proportion of migrants of the same ethnicity among the population in a region, which could be used as yet an alternative definition. This proportion could be interpreted as an indicator for the number of possible interactions with same-ethnicity individuals in random encounters in the region. We believe this measure to be of lesser importance to the decision maker, not only because regional migrant networks tend to be spatially concentrated even within the region[20] which makes random encounters less important, but also because migrants are more likely to value the highest possible number of interactions with same-ethnicity individuals rather than the probability of a random encounter with someone from the same country of origin.

[Table 5 about here.]

Table 5 shows the results of random parameters and conditional logit regressions of location choice. The main results of section 4 are unaltered by this change in the definition of the network variable: ethnic networks both in the same region as well as in neighboring regions (both within the country and across borders) significantly affect migrants' choice of the region of residence. The odds of choosing a region are 33.2 % larger if the local ethnic network increases by 1,000 individuals, and the effect of network size is again positive for all migrants. This also holds true for the effects of networks in neighboring regions which are, however, of limited importance in this specification: increasing the network size in a neighboring region by 1,000 individuals increases the odds of choosing a region by only 0.2 %. Networks in the rest of the country also play a role for the location decision, the effect is, however, rather small and even negative for about 63 % of all migrants. Compared to the estimation results of table 2 the proportion of individuals for which this coefficient is positive decreases by about 40 p. p. in reaction to the change in network definition. As before, ethnic networks in neighboring regions of other countries affect the location decision positively. In contrast to the relative network size regression (table 2), the estimated coefficients are slightly larger than for within-country neighbors, but the absolute network size estimation shows a higher degree of heterogeneity in these parameters. Again, the empirical model shows evidence in support of the optimal network size hypothesis because of the negative parameter of the squared local network.

[20] E. g., ethnic enclaves such as the Chinatowns in U. S. cities like San Francisco or New York or in European cities like Liverpool and London are well defined within a few city blocks.

A direct comparison of the random parameters and conditional logit regressions again shows that imposing fixed parameters on the empirical model would lead to severe biases in parameters with a high degree of heterogeneity, such as the capital, common border, and colony dummies, but also in the absolute network size variable.

5.2 Alternative definition of neighboring regions

In the previous regressions, spatial ethnic networks were defined by summing up the network size in neighboring NUTS-2 regions. This ignores that region size differs across countries. Although the "Nomenclature des Unités Territoriales Statistiques" (NUTS) should ensure at least some comparability across regions in the European Union, NUTS-2 regions across Europe are quite heterogeneous. While continental France is more than 1.5 times the size of Germany, there are 39 German and only 22 (continental) French NUTS-2 regions. In another example, while the Region Övre Norrland (NUTS-2 code: SE33) had an area of about 165,300 km^2 and about 3.3 inhabitants per km^2 in 2007, the region Bruxelles-Capitale (NUTS-2 code: BE10) had 161 km^2 and about 6,500 inhabitants per km^2 (in 2007) according to Eurostat data. This of course implies that the availability of a network in neighboring regions may differ with region size.

Therefore we also estimate the model defining neighboring regions by their distance to the migrant's region of residence. Ethnic networks in other regions are considered only if the geographical center of the neighboring region is within a radius ρ of 0–100 or 101–200 kilometers (as the crow flies) from the geographical center of the region of residence. Networks in the rest of the country (outside the 200 kilometer radius) are included as an additional regressor. As shown by the summary statistics in table 6, on average 4.0 % of an ethnic network (outside the region of residence) can be found within a 100 kilometer radius. The average number of those living in regions within 101–200 kilometers is about 5.1 %. As before, we differentiate between ethnic networks within the host country and networks in neighboring countries. Networks in neighboring regions in other countries are considered when they are within a radius ρ' of 0–100 and 101-200 kilometers. As before, the size of the local ethnic network in the region of residence as well as its squared value are included which allows a direct comparison with the results in table 2.

[Table 6 about here.]

Despite the change in the definition of neighboring regions, the main conclusions are unaltered (see table 7): even if networks in neighboring regions are considered only if they are within a given distance to the region of residence, they still affect location choice positively for all migrants. Networks within a 100 kilometer radius exert a larger influence than networks within a 200 kilometer radius or networks in the rest of the country. Again, the largest effect can be found for ethnic networks in the region chosen by the migrant. As before, this effect is decreasing in network size for most migrants, and only for 10 % the squared network variable has a positive coefficient. The estimated network parameter is slightly larger in this regression, but close to the original parameter of table 2.

[Table 7 about here.]

In contrast to the previous regressions, networks in neighboring regions of other countries within a 100 kilometer radius affect location choice negatively for 84 % of all migrants although the coefficient of first neighbors in other countries was significantly positive in the other regressions and networks in regions of other countries within in a 100–200 kilometer radius exert a positive influence on the probability of choosing a specific region. This can be explained by the differences in coverage between the two definitions: Each region has, on average, 0.53 neighboring regions in other countries, but only 0.29 regions in other countries are within a 100 kilometer radius.[21] Overall, there are only 21 regions where the closest region in another country is less than 100 kilometers away and actually hosts an ethnic network. The majority of these regions (16) are in Belgium and the Netherlands, and it is therefore likely that the difference to the main regression arises from the specifics of these countries or the ethnic groups living in these countries.

5.3 Migrants vs. refugees

Because we cannot distinguish between migrants and refugees in the EU-LFS data, our results might partly be driven by the differences in the location decisions of these groups. For example, in some European countries (especially Sweden, Denmark and the Netherlands) spatial dispersion policies are or have been in place which restrict(ed) the freedom of movement of refugees. Therefore, we also test the robustness of our results by estimating the model on subsets of origin countries defined by the level of development. We assume that individuals born in countries with low levels of development are more likely to be refugees who cannot freely choose their residence location.Migrants from these countries are therefore excluded from the regression.

[21] These differences are also substantial within countries: While each region has on average 3.52 (first) neighbors within the same country, only 1.42 (out of the first neighbors) are within a 100 kilometer radius, 1.30 are within 100–200 kilometers, and 0.80 are more than 200 kilometers away from the region of residence.

We use the United Nation's Human Development Index (HDI) to define the sending country's level of development.[22] The HDI ranges between zero and one, and countries are classified into one of the following four categories: very high human develoment ($0.9 \leq$ HDI ≤ 1), high human development ($0.8 \leq$ HDI < 0.9), medium human development ($0.5 \leq$ HDI < 0.8) and low human development ($0 \leq$ HDI < 0.5).

[Table 8 about here.]

To distinguish refugees from migrants, we estimate the model only for individuals born in countries with at least medium human development. The results in table 8 show that excluding those most likely to be refugees increases the size of the estimated parameter for local networks only slightly from 0.376 to 0.427. At the mean estimated parameter value, the odds of moving to a region are 53.3 % higher if the size of the ethnic network increases by 1 p. p. if refugees are excluded, an increase of 7.7 p. p. compared to the regression for all countries (table 2). The other estimated parameters are hardly affected by excluding refugees. Therefore, the main conclusions of our preferred specification remain intact.

5.4 Educational and retiree migration

Finally, to focus specifically on the effects of networks on labor migration the model was also estimated on a restricted sample including only migrants between 25 and 54 years of age. Both the location choice of younger migrants as well as the location choice of older migrants may be driven by characteristics not related to the labor market. For example, younger individuals moving abroad to study will choose their location based on education opportunities, and not based on regional labor market characteristics. The location choice of retired individuals, on the other hand, may be driven by factors such as climatic conditions, as evidenced not only by retiree migration to Florida but also by international retiree migration to mediterranean countries in Europe (see Warnes, 2009, for a recent review).[23]

[Table 9 about here.]

[22] The 2009 edition of the Human Development Report is used, which reports the HDI based on 2007 figures, see UNDP (2009) for details.

[23] The sample is restricted to migrants younger than 55 years of age because older cohorts would already contain a large number of retirees. According to 2006 data from the European Labour Force Survey (Eurostat, 2008), the average age at which employed persons started receiving a retirement pension in the countries considered ranged from 54.5 years in France to 61.7 years in Denmark.

The regression on the restricted sample shows that the main conclusions of our analysis are practically unchanged (9). As before, both local ethnic networks as well as ethnic networks in neighboring regions have a significantly positive effect on location choice of working-age migrants to Europe. Furthermore, quantitative differences between the estimated parameters are rather small, supporting the results of our main specification (see table 2).[24]

6 Summary

This paper analyzes the effect of ethnic networks on the location decision of migrants who moved to the EU between 1998 and 2007 using data from the European Labour Force Survey. Using a random parameters logit specification we find a substantially positive effect of ethnic networks on the location decision of migrants, providing strong evidence for ethnic clustering of migrants among European regions. The random parameters specification, which allows for individual heterogeneity in utility functions, shows that there are substantial variations in taste across individuals. The effect of ethnic networks in the same region is, however, positive for all individuals. We also find evidence of spatial spillovers in the effect of ethnic networks: ethnic networks in neighboring regions (both in the same country as well as across the border) and networks in the rest of the country significantly help to explain migrants' choice of target regions. The positive effects of ethnic networks thus extends beyond regional and national borders. Additional estimations using different network and neighborhood definitions as well as on subsamples of the data confirm the robustness of our findings.

We also find a sizable compensating variation (willingness to pay), especially for regions where only few previous migrants from the same country of origin are located. Individuals would require a compensation of about € 21,600 in order to consider a region without an ethnic network as attractive as an otherwise equivalent region with an ethnic network size of 1 %.Considering that the average annual income per employee among the 158 regions used in the analysis is about € 27,300, ethnic networks thus play a very important role in the location decision and regions without an ethnic network are highly unattractive.

[24] If migration decisions are made at the household level, Mincer (1978) showed that women are more likely to be "tied movers" while men more often are the "primary movers" of the household. If this is the case or if female household members move later than male household members (for example because family reunion in the host country is not immediately possible after migration), a female migrant's location choice may depend only on the choice made by her partner and not on other factors such as ethnic networks, which may affect the estimated parameters. But as shown by a regression on the restricted sample of male migrants aged 25 to 54 (see table A1 in the appendix) excluding females from the sample has only small quantitative effects on most estimated parameters, so the main conclusions of section 4 remain unchanged.

At the average network size of about 6.65 %, the average compensating variation is about € 9,500. There is, however, a considerable heterogeneity in the compensating variation across individuals and the compensating variation for networks in neighboring regions is considerably smaller than for networks within a region. There is also a substantial border effect for network externalities. Our results therefore show that ethnic networks in neighboring regions matter, but that the importance of networks decreases with distance to the region of residence, and that national borders reduce the positive effect of ethnic networks in neighboring regions.

We also find evidence for the optimal network size hypothesis. For most migrants, the positive effect of ethnic networks is decreasing in network size, as indicated by a negative squared network parameter. Based on individual level coefficients, the effect of the ethnic network becomes negative at a network size of 16.6 % on average. But only 0.8 % of all migrants live in a region where the actual network size exceeds the optimal network size. Although there is evidence for an optimal network size, it is rather a hypothetical construct and only few ethnic networks are actually close to the optimal level.

Although the qualitative conclusions concerning most estimated parameters are similar to those of the computationally simpler conditional logit model, our finding of spatial spillovers in the effect of ethnic networks shows that the conditional logit's independence from irrelevant alternatives (IIA) property is violated. Furthermore, the significant standard deviations of the random parameters show that the limitations imposed by the conditional logit on the individual parameters are too strict. We therefore conclude that the random parameters logit is superior to the conditional logit in the analysis of the location decision of migrants and that there can be considerable differences between conditional and random parameters estimates if there is a high degree of heterogeneity in the population.

Some policy conclusions can be drawn from the results of our analysis. Our results point to a strong "lock-in effect" of the ethnic structure of migration. The current regional ethnic structure of migration in part determines the future regional pattern of ethnic migration. This implies that the heterogeneous use of restrictions on the movement of labor among the EU-15 countries during the transitional period will have effects on the long-term migration patterns from the 8 member states which joined the EU in 2004. But regional concentrations of migrants of the same ethnicity can be detrimental to integration measures and foster the evolution of parallel societies. However, spatial dispersion policies (as employed for example in Sweden) which aim at breaking up regional patterns of ethnic migration will lead to a substantial welfare loss for migrants, which must be considered in the evaluation of such policies.

There is still scope for future extensions. First, the data set currently does not allow us to distinguish migrants from refugees, which might have completely different location patterns. Second, there may be differences according to education level of migrants. E. g., highly skilled migrants may avoid regions with large concentrations of low-skill migrants of the same ethnicity to escape statistical discrimination (cf. Stark, 1994). Third, it could be interesting to analyze the substitution patterns between regions based on the random parameters logit model. Analyzing these substitution patterns could, for example, shed light on the effects of changes in economic conditions (or migration policy) in one country (or region) on migration to all other countries (or regions) and thus provide us with an important tool to forecast future migration patterns based on past migration.

References

Åslund, O. (2005): "Now and forever? Initial and subsequent location choices of immigrants," *Regional Science and Urban Economics*, 35, 141–165.

Andersson, P. and E. Wadensjö (2007): "The employees of native and immigrant self-employed," IZA Discussion Paper 3147, Institute for the Study of Labour (IZA), Bonn.

Anselin, L. (2006): "Spatial Econometrics," in *Palgrave Handbook of Econometrics*, ed. by T. C. Mills and K. Patterson, New York: Palgrave Macmillan, vol. 1, chap. 26, 901–969.

Bartel, A. P. (1989): "Where do the new U. S. immigrants live?" *Journal of Labor Economics*, 7, 371–391.

Bauer, T., G. S. Epstein, and I. N. Gang (2000): "What are migration networks?" IZA Discussion Paper 200, Institute for the Study of Labour (IZA), Bonn.

——— (2002): "Herd effects of migration networks? The location choice of Mexican immigrants in the U. S." IZA Discussion Paper 551, Institute for the Study of Labour (IZA), Bonn.

——— (2005): "Enclaves, language and the location choice of migrants," *Journal of Population Economics*, 18, 649–662.

Bhat, C. R. (2001): "Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model," *Transportation Research B*, 35, 677–693.

Borjas, G. J. (1992): "Ethnic capital and intergenerational mobility," *The Quarterly Journal of Economics*, 107, 123–150.

——— (1995): "Ethnicity, neighborhoods, and human-capital externalities," *American Economic Review*, 85, 365–390.

Boyd, J. H. and R. E. Mellman (1980): "The effect of fuel economy standards on the U.S. automotive market: An hedonic demand analysis," *Transportation Research A*, 14, 367–378.

Cappellari, L. and S. P. Jenkins (2006): "Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation," *The Stata Journal*, 6, 156–189.

Cardak, B. A. and J. T. McDonald (2004): "Neighbourhood effects, preference heterogeneity and immigrant educational attainment," *Applied Economics*, 36, 559–572.

Cardell, N. S. and F. C. Dunbar (1980): "Measuring the societal impacts of automobile downsizing," *Transportation Research A*, 14, 423–434.

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

Chiswick, B. and P. W. Miller (2005): "Do enclaves matter in immigrant adjustment?" *City & Community*, 4, 5–35.

Christiadi and B. Cushing (2008): "The joint choice of an individual's occupation and destination," *Journal of Regional Science*, 48, 893–919.

Cutler, D. M. and E. L. Glaeser (1997): "Are ghettos good or bad?" *The Quarterly Journal of Economics*, 112, 827–872.

Dahlberg, M. and M. Eklöf (2003): "Relaxing the IIA assumption in locational choice models: a comparison between conditional logit, mixed logit, and multinomial probit models," Working Paper 2003:9, Department of Economics, Uppsala University, Uppsala.

Damm, A. P. (2009a): "Determinants of recent immigrants' location choices: quasi-experimental evidence," *Journal of Population Economics*, 22, 145–174.

——— (2009b): "Ethnic enclaves and immigrant labor market outcomes: quasi-experimental evidence," *Journal of Labor Economics*, 27, 281–314.

Davies, P. S., M. J. Greenwood, and H. Li (2001): "A conditional logit approach to U. S. state-to-state migration," *Journal of Regional Science*, 41, 337–360.

Drukker, D. M. and R. Gates (2006): "Generating Halton sequences using Mata," *The Stata Journal*, 6, 214–228.

Edin, P.-A., P. Fredriksson, and O. Åslund (2001): "Ethnic enclaves and the economic success of immigrants – evidence from a natural experiment," CEPR Working Paper 2729, Centre for Economic Policy Research, London.

Epstein, G. (2002): "Informational cascades and decision to migrate," IZA Discussion Paper 445, Institute for the Study of Labour (IZA), Bonn.

Eurostat (2008): *Transition from Work into Retirement*, Luxembourg: Office for Official Publications of the European Communities.

Geis, W., S. Uebelmesser, and M. Werding (2008): "How do migrants choose their destination country? An analysis of institutional determinants," CESifo Working Paper 2506, CESifo, Munich.

Gottlieb, P. D. and G. Joseph (2006): "College-to-work migration of technology graduates and holders of doctorates within the United States," *Journal of Regional Science*, 46, 627–659.

Grogger, J. and G. Hanson (2008): "Income maximization and the selection and sorting of international migrants," NBER Working Paper 13821, National Bureau of Economic Research, Cambridge, MA.

Gross, D. M. and N. Schmitt (2003): "The role of cultural clustering in attracting new immigrants," *Journal of Regional Science*, 43, 295–318.

Halton, J. H. (1960): "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, 2, 84–90.

Heitmueller, A. (2006): "Coordination failures in network migration," *The Manchester School*, 74, 701–710.

Hensher, D. A. (2001): "The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications," *Transportation*, 28, 101–118.

Hensher, D. A. and W. H. Greene (2003): "The mixed logit model: the state of practice," *Transportation*, 30, 133–176.

Hole, A. R. (2007): "Fitting mixed logit models by using maximum simulated likelihood," *The Stata Journal*, 7, 388–401.

Jaeger, D. A. (2007): "Green Cards and the location choices of immigrants in the United States," *Research in Labor Economics*, 27, 131–183.

Lazear, E. P. (1999): "Culture and language," *Journal of Political Economy*, 107, 95–126.

Marschak, J. (1960): "Binary choice constraints on random utility indications," in *Standord Symposium on Mathematical Methods in the Social Sciences*, ed. by K. Arrow, Stanford: Stanford University Press, 312–322.

Massy, D., J. Arango, G. Hugo, A. Kouaouci, A. Pellegriono, and J. Taylor (1993): "Theories of international migration: a review and appraisal," *Population and Development Review*, 19, 431–466.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, chap. 4, 105–142.

McFadden, D. and K. E. Train (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447–470.

Mincer, J. (1978): "Family migration decisions," *Journal of Political Economy*, 86, 749–773.

Munshi, K. (2003): "Networks in the modern economy: Mexican migrants in the U. S. labor market," *The Quarterly Journal of Economics*, 118, 549–599.

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

Portnov, B. A. (1999): "The effect of regional inequalities on migration: a comparative analysis of Israel and Japan," *International Migration*, 37, 587–615.

Revelt, D. and K. E. Train (1998): "Mixed logit with repeated choices: households' choices of appliance efficiency level," *The Review of Economics and Statistics*, 80, 647–657.

Ruud, P. A. (1996): "Approximation and simulation of the multinomial probit model: an analysis of covariance matrix estimation," Working paper, Department of Economics, University of California, Berkeley.

Sillano, M. and J. d. D. Ortúzar (2005): "Willingness-to-pay estimation with mixed logit models: some new evidence," *Environment and Planning A*, 37, 525–550.

Stark, O. (1994): "Patterns of Labor Migration when Workers Differ in Their Skills," in *Economic Aspects of International Immigration*, ed. by H. Giersch, Berlin: Springer Verlag.

Train, K. E. (1999): "Halton sequences for mixed logit," Working paper, Department of Economics, University of California, Berkeley.

——— (2009): *Discrete Choice Methods with Simulation*, New York: Cambridge University Press, second ed.

UNDP (2009): *Human Development Report 2009. Overcoming Barriers: Human Mobility and Development*, New York: United Nations Development Programme.

Warnes, T. A. M. (2009): "International retirement migration," in *International Handbook of Population Aging*, ed. by P. Uhlenberg, Netherlands: Springer, vol. 1 of *International Handbooks of Population*, 341–363.

Zavodny, M. (1999): "Determinants of recent immigrants' locational choice," *International Migration Review*, 33, 1014–1030.

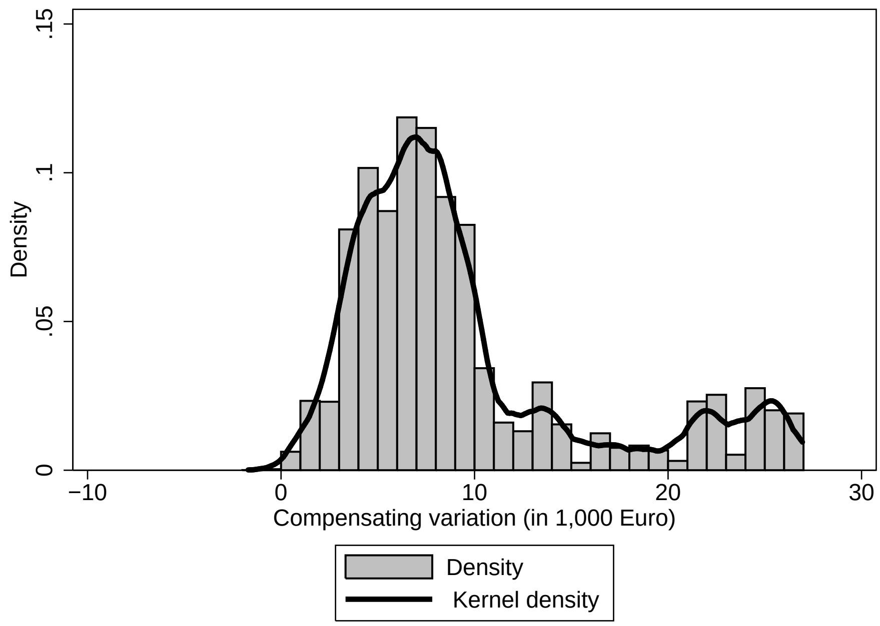

Figure 1: Distribution of compensating variation for a 1 percentage point variation in ethnic network size (in € 1,000). Histogram and kernel density estimate. Calculated from individual level parameters at mean network size. $N = 8,988,710$ observations. Source: European Labour Force Survey 2007.

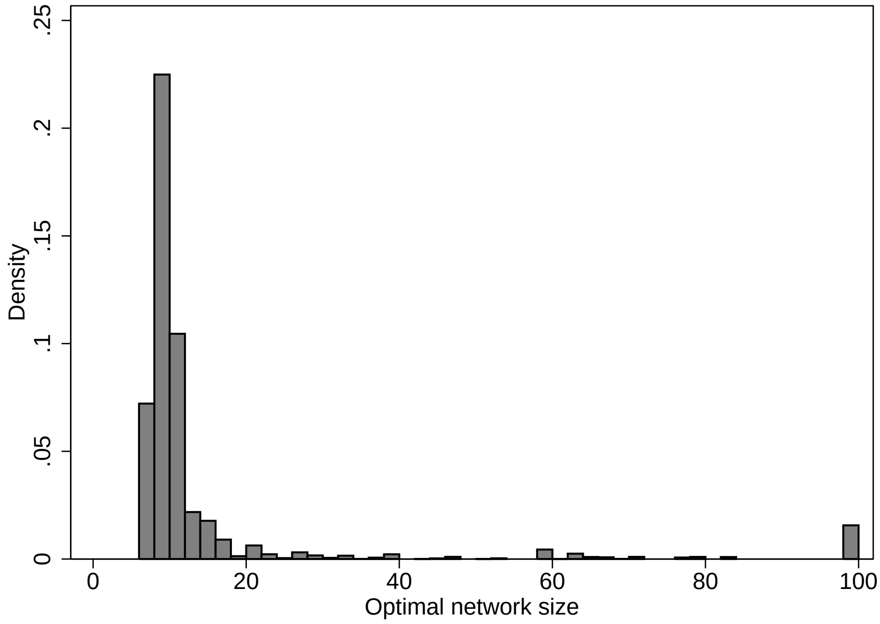

Figure 2: Distribution of optimal network size calculated from individual level parameters. $N = 8,375,514$ observations. Source: European Labour Force Survey 2007.

Variable	Mean	S. D.	Min.	Max.
Network_{js}	6.650	10.276	0.000	100.000
$\text{Network}_{js}^{N_1}$	7.449	9.841	0.000	100.000
$\text{Network}_{js}^{N_2}$	9.571	11.220	0.000	100.000
$\text{Network}_{js}^{N_C}$	14.442	17.334	0.000	100.000
$\text{Network}_{js}^{N_1'}$	0.364	1.568	0.000	32.338
$\text{Network}_{js}^{N_2'}$	1.310	3.570	0.000	51.233
Population (in 100,000)†	1.544	1.449	0.107	9.027
Region size (in 1,000 km^2)†	17.345	23.686	0.161	165.296
Unemployment rate (in %)†	7.290	3.743	2.286	20.186
Avg. income p. a. (in € 1,000)†	27.263	10.299	10.567	95.979
Capital (= 1)†	0.082	0.275	0.000	1.000
Distance (in 1,000 km)	4.697	3.641	0.055	18.981
Common border (= 1)	0.045	0.207	0.000	1.000
Common official language (= 1)	0.375	0.484	0.000	1.000
Colony after 1945 (= 1)	0.140	0.347	0.000	1.000

Table 1: Summary statistics of the independent variables. $^\dagger N = 158$ observations, all other variables: $N = 8,988,710$ observations. Source: European Labour Force Survey 2007, Eurostat, CEPII.

	RPL				CL	
Variable	Mean(β)	S.D.(β)	% $\beta > 0$	$e^{\text{Mean}(\beta)}$	β	e^{β}
Network_{js}	0.376***	0.014***	100.000	1.456***	0.127***	1.136***
	(0.001)	(0.000)		(0.001)	(0.000)	(0.000)
Network^2_{js}	−0.017***	0.013***	10.229	0.983***	−0.001***	0.999***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_1}_{js}$	0.051***	0.001***	100.000	1.053***	0.049***	1.051***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_2}_{js}$	0.036***	0.004***	100.000	1.037***	0.031***	1.031***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_C}_{js}$	0.025***	0.034***	76.856	1.025***	0.026***	1.027***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N'_1}_{js}$	0.034***	0.002***	100.000	1.035***	0.043***	1.044***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N'_2}_{js}$	0.012***	0.005***	98.720	1.012***	0.014***	1.014***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Population (in 100,000)	0.290***	0.001***	100.000	1.336***	0.324***	1.383***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Region size (in 1,000 km^2)	−0.005***			0.995***	−0.004***	0.996***
	(0.000)			(0.000)	(0.000)	(0.000)
Unemployment rate (in %)	−0.022***	0.005***	0.001	0.978***	−0.019***	0.981***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Avg. income p. a. (in € 1,000)	0.016***	0.001**	100.000	1.016***	0.017***	1.018***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Distance (in 1,000 km)	−0.203***	0.073***	0.280	0.816***	−0.412***	0.663***
	(0.001)	(0.005)		(0.001)	(0.001)	(0.001)
Distance (in 1,000 km)^2	0.012***	0.005***	99.523	1.012***	0.023***	1.023***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Capital (= 1)	−1.731***	3.297***	29.972	0.177***	−0.004***	0.996***
	(0.008)	(0.010)		(0.001)	(0.001)	(0.001)
Common border (= 1)	−0.264***	2.716***	46.135	0.768***	0.524***	1.689***
	(0.007)	(0.011)		(0.005)	(0.003)	(0.005)
Common official language (= 1)	1.088***	1.587***	75.355	2.968***	0.863***	2.369***
	(0.002)	(0.005)		(0.007)	(0.002)	(0.004)
Colony after 1945 (= 1)	−1.572***	1.280***	10.979	0.208***	−1.036***	0.355***
	(0.004)	(0.009)		(0.001)	(0.003)	(0.001)
Observations	8,988,710				8,988,710	

Table 2: Random parameters logit (RPL) and conditional logit (CL) regressions of location choice using relative network size. Germany and Ireland not included. Receiving country fixed effects reported in table A2. Standard errors in parentheses. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. RPL log likelihood simulated using 500 Halton draws. Source: European Labour Force Survey 2007, Eurostat, CEPII.

Variable	Mean	S. D.	Min.	Max.
$\mathrm{CV}_k(\mathrm{Network}_{js} = 1)$	21.612	0.982	19.734	24.290
$\mathrm{CV}_k(\mathrm{Network}_{js} = 5)$	13.010	4.809	4.564	26.170
$\mathrm{CV}_k(\mathrm{Network}_{js} = 10)$	2.257	9.594	-14.415	28.519
$\mathrm{CV}_k(\mathrm{Network}_{js} = 15)$	-8.495	14.379	-33.394	30.868
$\mathrm{CV}_k(\mathrm{Network}_{js} = 20)$	-19.248	19.164	-52.373	33.218
$\mathrm{CV}_k(\mathrm{Network}_{js} = 25)$	-30.000	23.949	-71.352	35.567
$\mathrm{CV}_k\left(\mathrm{Network}_{js}^{N_1}\right)$	3.245	0.005	3.184	3.285
$\mathrm{CV}_k\left(\mathrm{Network}_{js}^{N_2}\right)$	2.286	0.026	2.128	2.463
$\mathrm{CV}_k\left(\mathrm{Network}_{js}^{N_C}\right)$	1.558	0.677	-0.535	5.387
$\mathrm{CV}_k\left(\mathrm{Network}_{js}^{N_1'}\right)$	2.182	0.009	2.124	2.244
$\mathrm{CV}_k\left(\mathrm{Network}_{js}^{N_2'}\right)$	0.766	0.024	0.613	1.011

Table 3: Compensating variation (willingness to pay) for a 1 percentage point change in network size (in € 1,000). $N = 8,988,710$ observations. Source: European Labour Force Survey 2007.

Variable	Mean	S. D.	Min.	Max.
Absolute network$_{js}$ (in 1,000)	14.323	34.175	0.000	265.987
Absolute network$_{js}^{N_1}$ (in 1,000)	16.964	39.340	0.000	463.514
Absolute network$_{js}^{N_2}$ (in 1,000)	22.496	49.831	0.000	652.008
Absolute network$_{js}^{N_C}$ (in 1,000)	35.114	79.183	0.000	857.423
Absolute network$_{js}^{N_1'}$ (in 1,000)	2.564	13.955	0.000	365.762
Absolute network$_{js}^{N_2'}$ (in 1,000)	8.085	31.300	0.000	411.412

Table 4: Summary statistics, absolute network size variables. $N = 8,988,710$ observations. Source: European Labour Force Survey 2007.

Variable	RPL Mean(β)	RPL S.D.(β)	RPL % $\beta > 0$	RPL $e^{\text{Mean}(\beta)}$	CL β	CL e^{β}
Absolute Network$_{js}$ (in 1,000)	0.287***	0.035***	100.000	1.332***	0.045***	1.046***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Absolute Network$^2_{js}$ (in 1,000)	−0.006***	0.005***	11.132	0.994***	−0.000***	1.000***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Absolute Network$^{N_1}_{js}$ (in 1,000)	0.002***	0.000***	100.000	1.002***	0.003***	1.003***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Absolute Network$^{N_2}_{js}$ (in 1,000)	0.001***	0.000***	100.000	1.001***	0.001***	1.001***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Absolute Network$^{N_C}_{js}$ (in 1,000)	−0.001***	0.002***	36.764	0.999***	0.001***	1.001***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Absolute Network$^{N'_1}_{js}$ (in 1,000)	0.005***	0.002***	99.307	1.005***	0.002***	1.002***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Absolute Network$^{N'_2}_{js}$ (in 1,000)	0.002***	0.002***	80.126	1.002***	0.001***	1.001***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Population (in 100,000)	0.250***	0.003***	100.000	1.284***	0.357***	1.429***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Region size (in 1,000 km^2)	−0.001***			0.999***	−0.004***	0.996***
	(0.000)			(0.000)	(0.000)	(0.000)
Unemployment rate (in %)	−0.037***	0.035***	14.530	0.964***	−0.037***	0.964***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Avg. income p. a. (in € 1,000)	0.014***	0.014***	83.971	1.014***	0.021***	1.021***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Distance (in 1,000 km)	−0.474***	0.007**	0.000	0.622***	−0.610***	0.544***
	(0.001)	(0.003)		(0.001)	(0.001)	(0.001)
Distance (in 1,000 km)^2	0.021***	0.000***	100.000	1.021***	0.022***	1.022***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Capital (= 1)	−5.436***	8.016***	24.886	0.004***	0.105***	1.111***
	(0.025)	(0.033)		(0.000)	(0.001)	(0.001)
Common border (= 1)	0.464***	1.709***	60.697	1.590***	0.958***	2.608***
	(0.005)	(0.008)		(0.008)	(0.003)	(0.008)
Common official language (= 1)	1.689***	0.080***	100.000	5.416***	1.784***	5.955***
	(0.002)	(0.004)		(0.010)	(0.001)	(0.009)
Colony after 1945 (= 1)	−0.243***	1.059***	40.917	0.784***	0.066***	1.068***
	(0.004)	(0.007)		(0.003)	(0.003)	(0.003)
Observations	8,988,710				8,988,710	

Table 5: Random parameters logit (RPL) and conditional logit (CL) regressions of location choice using absolute network size. Germany and Ireland not included. Receiving country fixed effects reported in table A2. Standard errors in parentheses. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. RPL log likelihood simulated using 500 Halton draws. Source: European Labour Force Survey 2007, Eurostat, CEPII.

Variable	Mean	S. D.	Min.	Max.
$\text{Network}_j^{\rho\leq 100}$	4.008	9.965	0.000	97.181
$\text{Network}_j^{100<\rho\leq 200}$	5.106	9.777	0.000	100.000
$\text{Network}_j^{\rho>200}$	21.953	21.113	0.000	100.000
$\text{Network}_j^{\rho'\leq 100}$	0.061	0.857	0.000	60.556
$\text{Network}_j^{100<\rho'\leq 200}$	0.469	2.443	0.000	89.838

Table 6: Summary statistics, networks in neighboring regions defined by distance from region of residence. $N = 8,988,710$ observations. Source: European Labour Force Survey 2007.

Variable	RPL Mean(β)	RPL S.D.(β)	RPL % $\beta > 0$	RPL $e^{\text{Mean}(\beta)}$	CL β	CL e^{β}
Network_{js}	0.450***	0.001***	100.000	1.568***	0.135***	1.145***
	(0.001)	(0.000)		(0.001)	(0.000)	(0.000)
Network_{js}^2	−0.024***	0.019***	10.017	0.976***	−0.001***	0.999***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_j^{\rho \leq 100}$	0.041***	0.005***	100.000	1.042***	0.042***	1.043***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_j^{100<\rho \leq 200}$	0.036***	0.003***	100.000	1.037***	0.035***	1.035***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_j^{\rho > 200}$	0.033***	0.001***	100.000	1.033***	0.030***	1.031***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_j^{\rho' \leq 100}$	−0.099***	0.098***	15.750	0.906***	0.005***	1.005***
	(0.002)	(0.001)		(0.001)	(0.001)	(0.001)
$\text{Network}_j^{100<\rho' \leq 200}$	0.023***	0.001	100.000	1.024***	0.028***	1.029***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Population (in 100,000)	0.278***	0.006***	100.000	1.321***	0.317***	1.373***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Region size (in 1,000 km^2)	−0.002***			0.998***	−0.002***	0.998***
	(0.000)			(0.000)	(0.000)	(0.000)
Unemployment rate (in %)	−0.039***	0.010***	0.002	0.961***	−0.034***	0.967***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Avg. income p. a. (in € 1,000)	0.015***	0.010***	91.999	1.015***	0.019***	1.019***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Distance (in 1,000 km)	−0.212***	0.008*	0.000	0.809***	−0.463***	0.629***
	(0.001)	(0.004)		(0.001)	(0.001)	(0.001)
Distance (in 1,000 km)^2	0.011***	0.003***	99.936	1.011***	0.023***	1.023***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Capital (= 1)	−1.904***	3.384***	28.679	0.149***	−0.095***	0.909***
	(0.007)	(0.009)		(0.001)	(0.001)	(0.001)
Common border (= 1)	−0.138***	2.493***	47.795	0.871***	0.543***	1.721***
	(0.007)	(0.010)		(0.006)	(0.003)	(0.005)
Common official language (= 1)	1.081***	1.642***	74.480	2.946***	0.820***	2.270***
	(0.002)	(0.005)		(0.007)	(0.002)	(0.004)
Colony after 1945 (= 1)	−1.698***	1.587***	14.227	0.183***	−1.067***	0.344***
	(0.004)	(0.007)		(0.001)	(0.003)	(0.001)
Observations	8,988,710				8,988,710	

Table 7: Random parameters logit (RPL) and conditional logit (CL) regressions of location choice using relative network size. Germany and Ireland not included. Receiving country fixed effects reported in table A2. Standard errors in parentheses. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. RPL log likelihood simulated using 500 Halton draws. Source: European Labour Force Survey 2007, Eurostat, CEPII.

	RPL				CL	
Variable	Mean(β)	S. D.(β)	% $\beta > 0$	$e^{\text{Mean}(\beta)}$	β	e^{β}
Network_{js}	0.427***	0.042***	100.000	1.533***	0.132***	1.141***
	(0.001)	(0.001)		(0.001)	(0.000)	(0.000)
Network^2_{js}	−0.023***	0.018***	10.395	0.978***	−0.002***	0.998***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_1}_{js}$	0.052***	0.001***	100.000	1.054***	0.050***	1.051***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_2}_{js}$	0.036***	0.012***	99.881	1.037***	0.030***	1.031***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_C}_{js}$	0.025***	0.036***	76.131	1.026***	0.027***	1.027***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N'_1}_{js}$	0.029***	0.019***	93.831	1.029***	0.041***	1.042***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N'_2}_{js}$	0.011***	0.005***	98.060	1.011***	0.012***	1.013***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Population (in 100,000)	0.284***	0.003***	100.000	1.328***	0.317***	1.373***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Region size (in 1,000 km^2)	−0.004***			0.996***	−0.004***	0.996***
	(0.000)			(0.000)	(0.000)	(0.000)
Unemployment rate (in %)	−0.023***	0.006***	0.008	0.977***	−0.022***	0.979***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Avg. income p. a. (in € 1,000)	0.016***	0.005***	99.894	1.016***	0.018***	1.018***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Distance (in 1,000 km)	−0.168***	0.261***	25.948	0.845***	−0.401***	0.670***
	(0.001)	(0.004)		(0.001)	(0.001)	(0.001)
Distance (in 1,000 km)^2	0.009***	0.004***	99.258	1.009***	0.021***	1.021***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Capital (= 1)	−2.024***	3.636***	28.892	0.132***	0.022***	1.022***
	(0.008)	(0.011)		(0.001)	(0.001)	(0.001)
Common border (= 1)	−0.169***	2.561***	47.370	0.845***	0.538***	1.712***
	(0.007)	(0.010)		(0.006)	(0.003)	(0.005)
Common official language (= 1)	1.065***	1.554***	75.353	2.901***	0.834***	2.303***
	(0.003)	(0.006)		(0.007)	(0.002)	(0.004)
Colony after 1945 (= 1)	−1.792***	1.250***	7.584	0.167***	−1.115***	0.328***
	(0.005)	(0.009)		(0.001)	(0.003)	(0.001)
Observations		8,506,229			8,506,229	

Table 8: Random parameters logit (RPL) and conditional logit (CL) regressions of location choice using relative network size. Sample restricted to migrants from countries with at least "medium" human development (HDI ≥ 0.5, see UNDP, 2009). Germany and Ireland not included. Receiving country fixed effects reported in table A2. Standard errors in parentheses. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. RPL log likelihood simulated using 500 Halton draws. Source: European Labour Force Survey 2007, Eurostat, CEPII.

	RPL				CL	
Variable	Mean(β)	S. D.(β)	% $\beta > 0$	$e^{\text{Mean}(\beta)}$	β	e^{β}
Network_{js}	0.410***	0.026***	100.000	1.507***	0.124***	1.132***
	(0.001)	(0.001)		(0.001)	(0.000)	(0.000)
Network_{js}^2	−0.021***	0.016***	9.415	0.979***	−0.001***	0.999***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_{js}^{N_1}$	0.050***	0.002***	100.000	1.051***	0.050***	1.051***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_{js}^{N_2}$	0.033***	0.016***	98.249	1.034***	0.031***	1.031***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_{js}^{N_C}$	0.023***	0.034***	75.038	1.023***	0.026***	1.027***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}_{js}^{N_1'}$	0.022***	0.023***	82.607	1.022***	0.038***	1.039***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
$\text{Network}_{js}^{N_2'}$	0.016***	0.002***	100.000	1.017***	0.017***	1.017***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Population (in 100,000)	0.294***	0.000	100.000	1.341***	0.329***	1.390***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Region size (in 1,000 km^2)	−0.005***			0.995***	−0.005***	0.995***
	(0.000)			(0.000)	(0.000)	(0.000)
Unemployment rate (in %)	−0.021***	0.015***	8.821	0.979***	−0.017***	0.983***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Avg. income p. a. (in € 1,000)	0.015***	0.006***	99.265	1.015***	0.018***	1.018***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Distance (in 1,000 km)	−0.180***	0.058***	0.101	0.836***	−0.393***	0.675***
	(0.002)	(0.007)		(0.001)	(0.001)	(0.001)
Distance (in 1,000 km)^2	0.010***	0.002***	99.999	1.010***	0.022***	1.022***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Capital (= 1)	−1.668***	3.218***	30.214	0.189***	0.005***	1.005***
	(0.010)	(0.013)		(0.002)	(0.002)	(0.002)
Common border (= 1)	−0.170***	2.421***	47.207	0.844***	0.507***	1.661***
	(0.008)	(0.011)		(0.006)	(0.003)	(0.005)
Common official language (= 1)	1.087***	1.313***	79.604	2.965***	0.870***	2.386***
	(0.003)	(0.006)		(0.008)	(0.002)	(0.005)
Colony after 1945 (= 1)	−1.589***	1.484***	14.202	0.204***	−1.037***	0.355***
	(0.005)	(0.009)		(0.001)	(0.003)	(0.001)
Observations	6,732,410				6,732,410	

Table 9: Random parameters logit (RPL) and conditional logit (CL) regressions of location choice using relative network size. Sample restricted to migrants age 25 to 54. Germany and Ireland not included. Receiving country fixed effects reported in table A2. Standard errors in parentheses. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. RPL log likelihood simulated using 500 Halton draws. Source: European Labour Force Survey 2007, Eurostat, CEPII.

Variable	RPL Mean(β)	RPL S.D.(β)	RPL % $\beta > 0$	RPL $e^{\text{Mean}(\beta)}$	CL β	CL e^{β}
Network_{js}	0.341***	0.024***	100.000	1.406***	0.113***	1.120***
	(0.001)	(0.001)		(0.001)	(0.000)	(0.000)
Network^2_{js}	−0.015***	0.011***	9.261	0.985***	−0.001***	0.999***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_1}_{js}$	0.057***	0.006***	100.000	1.058***	0.053***	1.055***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_2}_{js}$	0.041***	0.007***	100.000	1.042***	0.035***	1.036***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N_C}_{js}$	0.030***	0.029***	84.604	1.030***	0.028***	1.029***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
$\text{Network}^{N'_1}_{js}$	−0.001**	0.066***	49.100	0.999**	0.036***	1.036***
	(0.001)	(0.001)		(0.001)	(0.000)	(0.000)
$\text{Network}^{N'_2}_{js}$	0.022***	0.008***	99.625	1.022***	0.022***	1.022***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Population (in 100,000)	0.324***	0.006***	100.000	1.383***	0.335***	1.398***
	(0.000)	(0.001)		(0.001)	(0.000)	(0.001)
Region size (in 1,000 km^2)	−0.008***			0.992***	−0.007***	0.993***
	(0.000)			(0.000)	(0.000)	(0.000)
Unemployment rate (in %)	−0.014***	0.001**	0.000	0.986***	−0.014***	0.986***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Avg. income p. a. (in € 1,000)	0.013***	0.015***	81.480	1.013***	0.017***	1.018***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Distance (in 1,000 km)	−0.205***	0.110***	3.134	0.815***	−0.385***	0.680***
	(0.002)	(0.007)		(0.002)	(0.002)	(0.001)
Distance (in 1,000 km)^2	0.010***	0.005***	97.856	1.010***	0.019***	1.019***
	(0.000)	(0.001)		(0.000)	(0.000)	(0.000)
Capital (= 1)	−1.755***	3.249***	29.455	0.173***	0.038***	1.038***
	(0.012)	(0.016)		(0.002)	(0.002)	(0.002)
Common border (= 1)	−0.658***	3.025***	41.389	0.518***	0.408***	1.505***
	(0.013)	(0.018)		(0.007)	(0.005)	(0.007)
Common official language (= 1)	1.097***	1.478***	77.093	2.995***	0.866***	2.378***
	(0.004)	(0.008)		(0.013)	(0.003)	(0.008)
Colony after 1945 (= 1)	−1.918***	1.271***	6.586	0.147***	−1.277***	0.279***
	(0.007)	(0.012)		(0.001)	(0.005)	(0.001)
Observations	3,177,240				3,177,240	

Table A1: Random parameters logit (RPL) and conditional logit (CL) regressions of location choice using relative network size. Sample restricted to male migrants age 25 to 54. Germany and Ireland not included. Receiving country fixed effects reported in table A2. Standard errors in parentheses. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. RPL log likelihood simulated using 500 Halton draws. Source: European Labour Force Survey 2007, Eurostat, CEPII.

	Table 2	Table 5	Table 7	Table 8	Table 9	Table A1
Belgium (=1)	−0.335***	−0.294***	−0.288***	−0.325***	−0.387***	−0.439***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)
Denmark (=1)	0.532***	0.404***	0.534***	0.285***	0.512***	0.410***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)
Spain (=1)	1.449***	1.744***	1.556***	1.530***	1.500***	1.602***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
Finland (=1)	−1.529***	−1.370***	−1.405***	−1.471***	−1.399***	−1.249***
	(0.011)	(0.009)	(0.011)	(0.011)	(0.013)	(0.021)
France (=1)	−0.964***	−0.824***	−0.838***	−0.838***	−0.958***	−0.991***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Greece (=1)	−0.609***	−0.334***	−0.512***	−0.555***	−0.562***	−0.325***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.007)
Italy (=1)	0.188***	0.657***	0.223***	0.258***	0.237***	0.128***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
Luxembourgh (=1)	−0.889***	−1.122***	−0.883***	−0.842***	−0.721***	−0.845***
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.017)
Netherlands (=1)	−0.405***	−0.082***	−0.395***	−0.317***	−0.295***	−0.389***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
Portugal (=1)	0.361***	0.437***	0.469***	0.433***	0.432***	0.659***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.007)
Sweden (=1)	0.129***	0.105***	0.141***	−0.006	0.190***	0.230***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)
United Kingdom (=1)	0.481***	0.763***	0.431***	0.498***	0.465***	0.638***
	(0.003)	−0.003	(0.003)	(0.003)	(0.003)	(0.005)

Table A2: Receiving country fixed effects, random parameters logit regressions. Base category: Austria. Germany and Ireland not included. Standard errors in parentheses. *** significant at 1 %, ** significant at 5 %, * significant at 10 %. RPL log likelihood simulated using 500 Halton draws. Source: European Labour Force Survey 2007, Eurostat, CEPII.