Network migration: do neighbouring regions matter?

Klaus Nowotny & Dieter Pennerstorfer

Network migration: do neighbouring regions matter?

Klaus Nowotny[a] and Dieter Pennerstorfer[b]

ABSTRACT
This paper analyses the role of the spatial structure of migrant networks in the location decision of migrants to the European Union at the regional level. Using a random parameters logit specification, a significant positive effect of migrant networks in neighbouring regions on migrants' location decisions is found. Although this spatial spillover effect is smaller than the effect of networks in the host regions, omitting to control for this spatial dependence results in a 40% overestimation of the effect of regional migrant networks on the location decision of newly arriving migrants.

KEYWORDS
network migration; location decisions; spatial heterogeneity; random parameters logit

INTRODUCTION

Previous research has shown not only that migrants' location choices can be explained by differences in economic opportunities (e.g., Davies, Greenwood, & Li, 2001), but also that migrants tend to settle where other migrants of the same ethnicity or from the same country of birth migrated previously. Since the seminal study on ethnic migrant concentration in the United States by Bartel (1989), several papers have formulated hypotheses explaining migrant concentrations theoretically (Carrington, Detragiache, & Vishwanath, 1996; Chiswick & Miller, 2005; Gross & Schmitt, 2003; Massey et al., 1993) and have analysed the importance of migrant networks (or 'diasporas'; Beine, Docquier, & Özden, 2011) for the location decision of migrants empirically (e.g., Åslund, 2005; Bauer, Epstein, & Gang, 2000; Beine et al., 2011; Beine, Docquier, & Özden, 2015; Damm, 2009; Gross & Schmitt, 2003; Pedersen, Pytlikova, & Smith, 2008; Zavodny, 1999; Zorlu & Mulder, 2008). All these studies find positive effects of networks on the location decisions of newly arriving migrants.

From a theoretical perspective these studies identify a number of channels through which networks increase the attractiveness of a region for newly arriving immigrants: networks can provide their members with ethnic goods such as food, clothing, social organizations, religious services, media (radio, newspapers etc.) or marriage markets (Chiswick & Miller, 2005). Similarly, newly arrived migrants can benefit from a better availability of information on housing or employment opportunities (Gross & Schmitt, 2003) or from job referrals by more established members of the network (Munshi, 2003). In addition, networks can provide help with the settlement process, decrease the perceived alienation in the host country (Bauer et al., 2000), provide financial assistance (Munshi, 2003) or reduce legal entry barriers via family-reunification programmes (Beine et al., 2015).

Empirically, nearly all studies find a positive impact of migrant networks within the same region on the region's attractiveness for new immigrants. But from a theoretical perspective, the effect of a network should not necessarily be confined to the – often administratively defined – region of residence. For instance, a region can be more attractive to immigrants if ethnic goods are produced in other nearby regions, and some ethnic goods might be provided only if the network size in all regions of a country is large enough (e.g., media). Migrants may choose to commute to work to a neighbouring region if job referrals are provided by networks in that region, and the provision of social or financial assistance may also not be limited by regional confines. A large network in neighbouring regions can also reduce

CONTACT
[a] (Corresponding author) klaus.nowotny@sbg.ac.at, klaus.nowotny@wifo.ac.at
Department for Social and Economic Sciences, University of Salzburg, Salzburg, Austria; and Austrian Institute of Economic Research WIFO, Vienna, Austria.
[b] dieter.pennerstorfer@jku.at
Austrian Institute of Economic Research WIFO, Vienna, Austria; Department of Economics, Vienna University of Economics and Business (WU), Vienna, Austria; and Department of Economics, Johannes Kepler University of Linz, Linz, Austria.

the risk associated with migrants' location decisions: if a target region is hit by a negative shock, migrants can move to an attractive region nearby at low (additional) migration costs.[1] Finally, even if regions were defined in functional rather than administrative terms, the markets where positive externalities of migrant networks can be expected (labour, housing or marriage markets, markets for ethnic goods or services etc.) may differ considerably in size, impeding to use a single 'local' network measure.

Thus, this paper contributes to the literature by accounting for the impact of migrant networks in neighbouring regions and other regions of the host country, next to the network in the region of residence, on migrants' location choices. Such impacts would provide indirect evidence for spillover effects of migrant networks across regional borders and/or for economies of scale in the production of ethnic goods, supporting the view that the impact of diasporas goes beyond the local level: regional developments in one region can benefit migrant communities in neighbouring regions, and 'communities on the move' can have far-ranging effects.

There is only limited empirical evidence on spillover effects of migrant networks so far. At a national level, Ruyssen and Rayp (2014) analyse bilateral migration flows between sub-Saharan African countries and find evidence that a change in the migrant stock between a pair of countries affects neighbouring countries of both the origin and the destination. Contrariwise, Jayet, Rayp, Ruyssen, and Ukrayinchuk (2016) take a regionally very disaggregated perspective and focus on municipalities of a single destination country (Belgium). They analyse migration from a number of sending countries separately and find positive effects for network size in both the target municipality and the neighbouring regions (of the target municipality) on the location choice of newly arriving immigrants for most sending countries. To the authors' knowledge, the present paper is the first contribution that empirically tests for the presence of spatial dependence in the effect of migrant networks on migrant's location choice at a regional level in a multi-country framework.

In addition, the empirical analysis uses a random parameters (mixed) logit framework (McFadden & Train, 2000) to model migrants' location choices, which relaxes the independence of irrelevant alternatives (IIA) assumption inherent in the conditional logit (CL) model and its Poisson equivalent usually applied in empirical research (e.g., Bertoli & Fernández-Huertas Moraga, 2015; Guimarães, Figueirdo, & Woodward, 2003, 2004; Schmidheiny & Brülhart, 2011). Furthermore, we also propose the random parameters model as an alternative way to deal with the issue of multilateral resistance to migration (Bertoli & Fernández-Huertas Moraga, 2013). The paper also contributes to the literature on migrant networks as there are only a few studies covering European countries, while most of the previous work focuses on the United States.[2]

The results, which are based on 2007 data from the European Union Labour Force Survey (EU-LFS), suggest that the probability of moving to a region depends not only on the local migrant network but also – albeit to a smaller extent – on the networks in adjacent regions and other regions of the country. When ignoring migrant networks in neighbouring regions, the effect of the local migrant network on the location choice of newly arriving migrants is overestimated by about 40%.

Analysing the link between migrant networks and the location choice of newly arriving migrants may therefore lead to biased results and incorrect policy conclusions if this form of spatial dependence is not accounted for. This is likely to be even more relevant if this question is analysed at a regionally more disaggregated level, and also has implications that go far beyond the location choice of immigrants. For instance, similar omitted variables problems are likely to arise when analysing the influence of regional migrant stocks on firm behaviour, as done, for example, by Joona and Wadensjö (2009) and Pennerstorfer (2016).

The paper is organized as follows. The empirical model used to estimate migrants' location choice is developed in the next section, which also gives a detailed description of the data used in the analysis. Estimation results are presented in the third section. Finally, the results and policy implications are discussed in the fourth section.

DATA AND ECONOMETRIC FRAMEWORK

Econometric method

To motivate the empirical analysis, migrants' location decisions are modelled based on a random utility maximization framework (Marschak, 1960): Each region r in the set of regions R yields a region-specific utility U_{kr}, and migrant k chooses region $s \in R$ if and only if $U_{ks} > U_{kr} \forall r \neq s$. Because the decision-maker's utility is unknown, observable characteristics of the regions X_{kr} are used to define the representative utility function $V_{kr} = V(X_{kr})$. If V_{kr} is linear in X_{kr}, the utility function U_{kr} can be written as:

$$U_{kr} = V_{kr} + \varepsilon_{kr} = \beta' X_{kr} + \varepsilon_{kr} \qquad (1)$$

where β is the vector of coefficients; and ε_{kr} is a random error term. Assuming that ε_{kr} is an independent and identically distributed (i.i.d.) extreme that follows an extreme value distribution, the probability that individual k chooses location s, P_{ks}, can be estimated by a CL model (McFadden, 1974; also Bartel, 1989; Bauer et al., 2000; Bauer, Epstein, & Gang, 2005, 2007; Christiadi & Cushing, 2008; Gottlieb & Joseph, 2006; Jaeger, 2007; Grogger & Hanson, 2011):

$$P_{ks} = \frac{\exp{(\beta' X_{ks})}}{\sum_{r=1}^{R} \exp{(\beta' X_{kr})}} \qquad (2)$$

Because the log-likelihood functions are similar up to a constant, the probability in equation (2) can also be estimated in a Poisson framework (Guimarães et al., 2003, 2004; Santos Silva & Tenreyro, 2006; Schmidheiny & Brülhart, 2011).

The CL as well as its Poisson equivalent imply IIA, which states that the ratio of choice probabilities 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. But IIA is violated if networks in neighbouring regions matter: the ratio of choice probabilities then no longer depends on the characteristics of s and t alone, but also on the characteristics (especially the networks; see below) of their neighbours. Including the network size in neighbouring regions among the regressors is thus also a test for IIA (also Train, 2009, p. 49).

This calls for a model that relaxes IIA. One step in this direction is a nested logit model, as applied by Jayet et al. (2016). However, a nested logit still imposes IIA within nests. Therefore, this paper uses the more flexible random parameters logit (RPL) that does not impose IIA (for an overview, see Hensher & Greene, 2003; McFadden & Train, 2000; and Train, 2009; for an application to migration research, see Gottlieb & Joseph, 2006). The RPL model can be derived from utility-maximizing behaviour by allowing the parameters of the characteristics X_{kr} in the representative utility function to vary over individuals:

$$U_{kr} = \beta'_k X_{kr} + \varepsilon_{kr} \qquad (3)$$

where β_k is a vector of coefficients for individual k representing k's preferences; and ε_{kr} is an i.i.d. error term. Thus, the utility function is heterogeneous across individuals and the coefficients in β_k are assumed to vary over decision-makers according to the density $f(\beta|\theta)$. This so-called 'mixing distribution' describes the distribution of the coefficients β conditional on the parameters θ. The econometric model in this paper follows Gottlieb and Joseph (2006) in assuming that the main coefficients in β (i.e., the effects of migrant networks; see below) are normally distributed. The parameters θ to be estimated for these coefficients are thus the mean and standard deviation of a normal distribution. All other coefficients are modelled as 'fixed' parameters, i.e., parameters whose standard deviation is restricted to zero (Hensher & Greene, 2003). Estimation of the RPL is based on maximum simulated likelihood (for details, see Appendix A in the supplemental data online).

Allowing the parameters to vary also absorbs individual heterogeneity that could lead to a correlation across choice alternatives, and thus to multilateral resistance to migration (Bertoli & Fernández-Huertas Moraga, 2013; also Beine, Bertoli, & Fernández-Huertas Moraga, 2016): As already noted by Bertoli and Fernández-Huertas Moraga (2015, p. 2), '[t]he assumption [...] that the vector of parameters β does not vary across individuals implies that any heterogeneity in the relationship between x_{ijk} and U_{ijk} ends up in ε_{ijk}, introducing a correlation in the stochastic component of utility across destinations'.[3] By relaxing this assumption and allowing the parameters to vary, the RPL can thus be considered an alternative way to deal with the issue of multilateral resistance.[4]

Migration data

This paper uses individual-level microdata from the 2007 EU-LFS to test the importance of the regional structure of migrant networks for the location decision of migrants in Europe. The EU-LFS is a large household survey conducted among approximately 1.2 million people in the EU on a quarterly basis (EUROSTAT, 2009, 2013). The annualized EU-LFS data used in this paper are calculated from averages of the quarterly data and thus cross-sectional in nature.

While EU-LFS data disseminated by EUROSTAT usually contain only aggregated information about the sending countries, the microdata available to the authors provide detailed information on migrants' countries of birth as well as the regions of residence at the NUTS-2 level, which allows the observation of migration stocks on a detailed place-to-place basis. In addition, the microdata distinguish between recent migrants (those who have moved during the last decade) and previous migrants (those who migrated more than a decade ago). It also contains survey weights, which give the number of individuals represented by each observation.

One limitation of the EU-LFS is that it only provides information about the stock of migrants who were living in the respective region at the time of the interview. Therefore, it is not possible to investigate migrants' initial location choice and subsequent mobility (Zorlu & Mulder, 2008) to analyse how the effect of migrant networks on the location choice of new migrants evolves over time, or to consider repeat and return migration.

The definition of migrant status is based on the country of birth: all individuals born outside their country of residence are considered migrants. As the aim of the paper is to identify the factors that determine the regional location decisions of migrants to the European Union, the focus is on individuals born outside the EU-27 (i.e., the EU member states as of 2007). Migration within the EU is not considered, and migrants who have moved from one EU country to another are not included in our sample. Because information on the country of birth is unavailable for both Germany and Ireland, only the 158 NUTS-2 regions in the other 13 EU-15 countries (Austria, Belgium, Denmark, Finland, France, Greece, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the UK; henceforth the EU-13) are considered as receiving regions.[5] Furthermore, the analysis is restricted to recent migrants who moved to the EU between 1998 and 2007, and who were between 25 and 64 years of age in 2007. The number of those who migrated more than a decade ago is used to calculate migration networks (see below).

The number of observations by country and in total are shown in Table 1, which gives the (unweighted) number of observations in the sample as well as the (weighted) number of migrants using the weights provided in the EU-LFS. The empirical analysis is based on 16,830 individual-level observations representing 5,417,300 recent migrants from 137 sending countries who moved to the EU-13 between 1998 and 2007.[6] For the full set of sending countries, see Appendix C in the supplemental data online.

Table 1. Number of recent migrants to EU-13 countries by receiving countries.

| Receiving country | Number of observations | | | |
| | Unweighted | | Weighted | |
	N	%	N	%
Austria	1410	8.38	149,600	2.76
Belgium	1354	8.05	129,500	2.39
Denmark	570	3.39	49,200	0.91
Finland	62	0.37	12,200	0.23
France	476	2.83	533,800	9.85
Greece	1057	6.28	143,100	2.64
Italy	5305	31.52	909,400	16.79
Luxembourg	222	1.32	6800	0.13
Netherlands	789	4.69	157,800	2.91
Portugal	644	3.83	120,300	2.22
Spain	1507	8.95	1,948,500	35.97
Sweden	1513	8.99	118,800	2.19
UK	1921	11.41	1,138,400	21.01
Total	16,830	100	5,417,400	100

Note: Weighted numbers are based on weights provided in the European Union Labour Force Survey (EU-LFS) and rounded to the nearest 100. The data sample includes only recent migrants who moved to the EU-13 between 1998 and 2007, and who were between 25 and 64 years of age in 2007.
Sources: EU-LFS (EUROSTAT, 2009, 2013) and authors' own calculations.

Migrant networks

The main variable of interest is the size of the migrant network, which is calculated using the number of migrants from the same country of birth who were living in the respective region in 2007 but had immigrated to their current country of residence before 1998.[7] The size of the network for a migrant from country j who moved to region s between 1998 and 2007 is defined as:

$$\text{Network}_{js}(s) = \ln(m_{js} + 1) \qquad (4)$$

where m_{js} is the stock of previous migrants from the same country of birth j who had moved to region s before 1998. This definition follows Bertoli and Fernández-Huertas Moraga (2015) and Ortega and Peri (2009, 2013) by adding 1 to m_{js} to avoid losing observations because the logarithm is not defined for $m_{js} = 0$. This is especially important for migrants from small sending countries (see below). The logarithmic specification reflects the assumption of a decreasing marginal utility of migrant networks, so that an increase in m_{js} has a smaller effect on the probability of choosing a specific region as the size of the network increases.

To test the hypothesis that migrant networks in neighbouring regions matter, the empirical analysis includes the size of the networks in neighbouring regions as an additional variable in the regression:

$$\text{Network}_{js}(N_s^1) = \ln\left(\sum_{n_s^1 \in N_s^1} m_{jn_s^1} + 1\right) \qquad (5)$$

where $N_s^1 \subset R$ is the set of regions in the country of residence sharing a border with region s (first-order neighbours). Furthermore, the empirical analysis includes the size of the networks in second-order neighbour regions N_s^2, i.e., regions in the country of residence that do not share a border with s but which border on neighbouring regions of s (except for s itself and the regions in N_s^1: $N_s^2 \subset R$, $\{s, N_s^1\} \cap N_s^2 = \emptyset$):

$$\text{Network}_{js}(N_s^2) = \ln\left(\sum_{n_s^2 \in N_s^2} m_{jn_s^2} + 1\right) \qquad (6)$$

Finally, if there are ethnic goods with strong economies of scale in production, the size of the migrant network in the rest of the host country can also affect the location decision. The log sum of migrant networks in the rest of the host country ($N_s^C \subset R$, $\{s, N_s^1, N_s^2\} \cap N_s^C = \emptyset$) therefore also enters the regression.

The sets in N_s^1, N_s^2 and N_s^C contain only regions within the same country because it can be expected that the effect of networks in neighbouring regions across the border differs from the effect of networks within a country. Yet, for border regions networks in neighbouring regions of other countries may be of importance, although it can be expected that they affect the location choice of migrants to a lesser extent, if at all: networks in neighbouring countries will not be able to help with immigration issues and bureaucratic structures because of national differences in migration regimes and procedures. Furthermore, labour 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. Neighbouring regions in other countries are thus included in the alternative sets $N_s^{1'}$ (first neighbours, $N_s^1 \cap N_s^{1'} = \emptyset$) and $N_s^{2'}$ (second neighbours, $N_s^2 \cap N_s^{2'} = \emptyset$). The log sums of the networks in $N_s^{1'}$ and $N_s^{2'}$ (again augmented by 1) are considered as additional variables in the regression. If significant, the coefficients can, in comparison with their within-country counterparts, provide information about border effects in network externalities.

Following the classification of spatial econometric models by LeSage and Pace (2009), this model can be described as a 'spatial lag of X' (SLX) model, with networks in neighbouring regions as spatially lagged explanatory variables.[8] By differentiating between migrant networks in the target region, first- and second-order neighbour regions, as well as the remaining regions of the country, we go one step further as do Beine et al. (2015), who distinguish between the network size at regional and national levels. To the best of the authors' knowledge, this is the first

Table 2. Summary statistics for the explanatory variables.

Variable	Mean	Standard deviation	Minimum	Maximum
$\text{Network}_{js}(s)$	1.207	2.584	0.000	12.491
$\text{Network}_{js}(N_s^1)$	2.208	3.377	0.000	13.047
$\text{Network}_{js}(N_s^{1'})$	0.476	1.726	0.000	12.810
$\text{Network}_{js}(N_s^2)$	2.590	3.588	0.000	13.388
$\text{Network}_{js}(N_s^{2'})$	1.218	2.698	0.000	12.927
$\text{Network}_{js}(N_s^C)$	3.038	3.887	0.000	13.891
Common language ($= 1$)	0.122	0.327	0.000	1.000
Colony after 1945 ($= 1$)	0.086	0.280	0.000	1.000
ln (Distance in 1000 km)	1.659	0.670	−2.389	2.992

Note: $N = 21,646$ sending country–receiving region dyads; N_s^1 = set of neighbouring regions within the host country; $N_s^{1'}$ = set of neighbouring regions in other countries; N_s^2 = set of second neighbour regions within the host country; $N_s^{2'}$ = set of second neighbour regions in other countries; N_s^C = set of all other regions in the host country.
Sources: EU-LFS (EUROSTAT, 2009, 2013), Mayer and Zignano (2011), and authors' own calculations.

paper explicitly to incorporate this form of spatial heterogeneity in the context of migrants' location choice.[9]

The econometric model specifies the parameters of all network variables as being normally distributed, and a mean and standard deviation will be estimated to capture the heterogeneity in the effect of networks on location decisions. This heterogeneity may, for example, arise if migrants from the same country of birth belong to different ethnic groups or because of individual differences in the reliance on ethnic networks.

Summary statistics of the network variables for all sending country-receiving region dyads are displayed in Table 2. For about 81% of all dyads the size of the network is zero. However, this share varies considerably across sending countries (from 23.4% for migrants from the United States to 99.4% for migrants from Tajikistan) and is, as expected, negatively correlated with the total number of migrants from a particular sending country ($\rho = -0.609$). Across all dyads, an average of 400 individuals from the same country of birth are living in the region of residence, with another 1700 in first-neighbour regions, 2400 in second-neighbour regions and 4700 in other regions of the same country. The average network size in surrounding regions in other countries is considerably smaller (around 200 in first-neighbour regions and 1000 in second-neighbour regions), as many regions do not border on regions of other countries.

Other explanatory variables

As control variables, V_{kr} first includes a set of region-specific dummies that capture the effects of variables that vary only over alternatives, but not over decision-makers (including, for example, regional wage or unemployment rates). The alternative-specific dummies also capture the average effect of all factors that are not in the model and ensure that the error term in (1) has zero mean (Train, 2009, p. 20).[10]

Second, country-pair specific attributes are added: these include a dummy variable for linguistic closeness taken from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII), which measures whether there

is a language spoken by at least 9% of the population in both the sending and the receiving countries ($= 1$; 0 otherwise) (Mayer & Zignano, 2011). A common language not only reduces migration costs (Pedersen et al., 2008), but can also raise the returns to skill in the host country (Grogger & Hanson, 2011). Since colonial ties can also affect the location choice of migrants, information on colonial relationships from the CEPII is also included, capturing whether two countries were in a colonial relationship after 1945 ($= 1$; 0 otherwise) (Mayer & Zignano, 2011).

Third, to proxy for the costs of migration (or the costs of visiting relatives at home), the distance (in kilometres, as the crow flies) between the capital of the sending country and the geographical centre of the region of residence is also included. To capture a possibly decreasing negative effect, this variable enters the regression in logarithmic terms.[11]

Representative utility V_{kr} is thus a linear function of receiving region-specific (dummy) variables, country-pair-specific variables (common language and colonial ties), as well as sending country–receiving region-specific variables (distance and migrant networks), which are assumed to determine the location choice of migrants. In contrast to the network variables, which are modelled as normally distributed parameters, the control variables are assumed to be fixed.[12]

ESTIMATION RESULTS

Table 3 shows the results of the RPL model of location choice estimated on the pooled sample of migrants from all 137 sending countries who moved to the above-mentioned 13 host countries between 1998 and 2007 (model 1).[13]

In addition to the estimated random parameters' means and standard deviations, Table 3 also shows the proportion of the parameters' PDF which is above zero. This gives the percentage of the sample for which the parameter is positive. If part of a coefficient's distribution is below zero, the variable constitutes an attractor for some, and a repellent for other individuals. Model 2 reports the coefficients

Table 3. Random parameters logit (RPL) and conditional logit (CL) regressions of location choice.

Variable	Model 1 RPL			Model 2 CL	Model 3 RPL			Model 4 CL
	Mean (β)	SD (β)	% $\beta > 0$	β	Mean (β)	SD (β)	% $\beta > 0$	β
Network$_{js}(s)$	0.279***	0.228***	88.946	0.185***	0.395***	0.281***	92.009	0.269***
	(0.019)	(0.019)		(0.009)	(0.018)	(0.017)		(0.009)
Network$_{js}(N_s^1)$	0.152***	0.153***	83.976	0.106***				
	(0.017)	(0.020)		(0.011)				
Network$_{js}(N_s^{1'})$	−0.019	0.144***	44.751	0.015				
	(0.017)	(0.023)		(0.010)				
Network$_{js}(N_s^2)$	0.127***	0.126***	84.326	0.086***				
	(0.015)	(0.020)		(0.010)				
Network$_{js}(N_s^{2'})$	0.011	0.020	70.884	0.013				
	(0.008)	(0.050)		(0.008)				
Network$_{js}(N_s^C)$	0.063***	0.074***	80.271	0.067***				
	(0.014)	(0.027)		(0.010)				
ln (Distance)	−0.416***			−0.425***	−0.807***			−0.857***
	(0.067)			(0.064)	(0.058)			(0.054)
Common language	1.027***			1.014***	1.473***			1.441***
	(0.060)			(0.057)	(0.060)			(0.058)
Colony after 1945	−0.512***			−0.277***	−0.074			0.172**
	(0.094)			(0.077)	(0.084)			(0.074)
Region-specific fixed effects		Yes		Yes		Yes		Yes
Observations		16,830		16,830		16,830		16,830
Log-likelihood		−18,605.649		−18,673.276		−18,928.161		−19,021.554

Note: All regressions include 157 region/alternative-specific fixed effects. Standard errors are given in parentheses. ***Significant at 1%; **significant at 5%; *significant at 10%. RPL log-likelihood was simulated using 100 Halton draws. N_s^1 = set of neighbouring regions within the host country; $N_s^{1'}$ = set of neighbouring regions in other countries; N_s^2 = set of second neighbour regions within the host country; $N_s^{2'}$ = set of second neighbour regions in other countries; N_s^C = set of all other regions in the host country. SD, standard deviation.
Sources: EU-LFS (EUROSTAT, 2009, 2013), Mayer and Zignano (2011), Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) and authors' own calculations.

of a CL regression using the same variables as model 1. Although the CL's IIA assumption is violated if the hypothesis of spatial spillovers in network effects is correct, the CL can still serve as an approximation to a model which relaxes this assumption (cf. Dahlberg & Eklöf, 2003). In addition, the comparison with the CL model highlights the error resulting from neglecting the heterogeneity across decision-makers. To compare the model including networks in neighbouring regions with a model that does not consider this type of spatial heterogeneity, models 3 and 4 report regression results of RPL and CL regressions that include only the migrant network in the region of residence. Note that all models include 157 receiving region fixed effects, which are not reported to save space.

The results of the RPL (model 1) show that, as expected, a large network of migrants from the same sending country living in the same region (Network$_{js}(s)$) increases the probability of new migrants choosing this particular region. However, it also shows that networks in neighbouring regions matter: the estimated probability of choosing a specific region also increases with the size of migrant networks in neighbouring regions of the same country, as documented by the positive and significant mean parameter estimates on Network$_{js}(N_s^1)$ and Network$_{js}(N_s^2)$. Even networks in other regions of the country (Network$_{js}(N_s^C)$) have a positive and statistically significant influence on migrants' location choice. The size of the effect decreases considerably with distance: the estimated mean parameter is only 54.5% of the effect of Network$_{js}(s)$ for first-neighbour regions, 45.5% for second-neighbour regions and 22.6% for the rest of the country. However, spatial spillover effects do not extend beyond national borders: both estimated mean parameters on Network$_{js}(N_s^{1'})$ and Network$_{js}(N_s^{2'})$ are not statistically significant.

The RPL model estimates both the mean and standard deviation of a random parameter. As Table 3 shows, the estimated standard deviations are sizeable and statistically significant for all network variables (except for migrant networks in second-neighbour regions of another country). This suggests a large degree of heterogeneity between individuals: despite the finding that recent migrants generally prefer regions with larger networks, there is a notable minority of about 11% who are deterred by a large migrant network in the respective region. One possible explanation for this is that some (especially highly skilled) migrants want to avoid the statistical discrimination they would experience in regions with large networks of low-skilled migrants.

The heterogeneous (and also sometimes negative) effect of migrant networks in first- and second-neighbour regions is also plausible: as the network in regions neighbouring on s grows, these regions might become more attractive than s itself. For some regions a large network in neighbouring regions may thus have a negative influence on the number of migrants. Although its estimated mean parameter is not statistically different from zero, the estimated standard deviation for the network in neighbouring regions of other countries is also statistically significant: The distribution of the parameter is thus centred on zero, but the

parameter nevertheless varies across the population. This kind of heterogeneity could not be identified by using, for example, the CL model, which supports the decision to estimate the model using RPL.

Turning to the other variables included in the regression, distance is found to have a negative effect on the location decision, while a common language increases the probability of choosing a specific region. The regression also shows that a past colonial relationship between the sending and receiving countries does not make a region more attractive to new migrants, which seems to be in contrast to results found by previous research (e.g., Grogger & Hanson, 2011; Ortega & Peri, 2009, 2013). One explanation for this is that a common colonial history is correlated with having a common language ($\rho = 0.533$), so that a large part of the effect of a shared history is already captured by the common language dummy. Indeed, while 68.9% of all sending country–receiving region dyads with a colonial relationship after 1945 share a common language, the same is true for only 6.9% of all dyads without a colonial history. Besides, in the receiving countries analysed here, immigration laws for former colonies were rather loose not only during the colonial relationship but also even up to the 1970s.[14] The stock of migrants from former colonies thus comprises people who migrated under preferential conditions. As immigrants from former colonies have mainly lost their preferential status, having a common colonial heritage is associated with a tightening of immigration laws over the last decades relative to regulations regarding immigration from other countries.[15]

Two comparisons are worth exploring. The first is a comparison of model 1 in Table 3 including the networks in neighbouring regions and model 3, which was also estimated by RPL but ignores the spatial structure of migrant networks. Generally, as the error terms in logit models are normalized and the coefficients are thus scaled by the standard deviation of the factors that are not included in the model (Train, 2009, pp. 40f.), it can be expected that parameter estimates become larger as more explanatory variables enter the model, and vice versa.[16] However, the estimated parameter on Network$_{js}(s)$ in model 3 is not smaller, but about 40% larger than in model 1. The most likely reason for this is an upward omitted variable bias in model 3 due to the positive effect of migrant networks in neighbouring regions on location decisions (see model 1) and the positive correlation between network sizes of neighbouring regions.[17] Ignoring the spatial structure of migrant networks thus leads to an overestimation of the effect of networks on location decisions.

The second comparison of interest is between an RPL (model 1) and a CL model that contains the same variables (model 2). Although the results of the RPL already show that the CL does not fit the data because it neglects important heterogeneity (as shown by the significant standard deviations of the estimated parameters) and because the IIA assumption is violated (as evidenced by the significant coefficients of networks in neighbouring regions), such a comparison can highlight the bias caused by imposing a CL structure on a data process characterized by a high

degree of heterogeneity. This comparison is thus especially interesting in light of the fact the CL specification is widely used, either directly or indirectly through a Poisson approximation (e.g., Beine et al., 2011; Bertoli & Fernández-Huertas Moraga, 2015; Guimarães et al., 2003, 2004; Schmidheiny & Brülhart, 2011).

Table 3 shows that although the qualitative conclusions from both models are rather similar, the quantitative differences are quite substantial: The parameter estimates on $\text{Network}_{js}(s)$, $\text{Network}_{js}(N_s^1)$ and $\text{Network}_{js}(N_s^2)$ are about one-third smaller in the CL than in the RPL. These differences are sizeable and also statistically significant. The CL thus underestimates the effect of migrant networks, and the evidence provided here does not support the hypothesis that CL can be used as an approximation to RPL.

CONCLUSIONS

This paper analyses the effect of the spatial structure of migrant networks on the location decision of migrants who moved to the EU between 1998 and 2007, using an RPL specification which allows for heterogeneous utility functions. It provides evidence of spatial spillovers in the effect of migrant networks: networks in neighbouring regions of the same country and in the rest of the country significantly help to explain migrants' choice of target regions. The positive effect of networks is thus not confined to regional borders: newly arrived migrants seem to benefit from networks in neighbouring regions as well. In line with the previous empirical literature, a substantially positive effect of migrant networks in the host region on the location decision of newly arriving migrants is also found, providing strong evidence of clustering of migrants by sending countries among European regions. Ignoring the spatial structure of migrant networks (i.e., omitting the network size of neighbouring regions), however, leads to a sizeable upward bias of the estimated effect of the network in a particular region on the migrants' location decision of about 40%.

The empirical evidence of spatial spillovers violates the IIA assumption, rendering the more conventional CL model invalid. This makes RPL the appropriate model for estimating location choice at the individual level when the characteristics of neighbouring regions affect the location decision. Comparing both methods shows that the results differ significantly and substantially between RPL and CL. Furthermore, the significant standard deviations in the RPL reveal substantial taste variations across individuals and show that the limitations imposed by CL on the individual parameters are too strict. It can therefore be concluded that RPL is superior to CL in the analysis of the location decision of migrants and that there are considerable differences between the models if there is a high degree of heterogeneity in the population and if spatial spillover effects exist. In addition, RPL can be considered an alternative way to deal with the issue of multilateral resistance.

The presence of regional spillover effects suggests that regional diasporas can be persistent if the spillover effects mitigate the need for spatial relocation. This implies that countries that employ spatial dispersion policies (e.g., for refugees) can expect a more balanced distribution of follow-up migrants. Future research (following, for example, the approach of Åslund, 2005) could thus test whether secondary migration of migrants subject to spatial dispersion policies (i.e., movements within the host country after initial placement) is less common among migrants from sending countries with larger networks in neighbouring regions. The results also suggest that if the aim of spatial dispersion policies is to prevent the concentration of migrants in a single region (mostly the capital), it should not be designed in such a way as to 'thin out' the migrant network across too many regions. If both the local network and the networks in neighbouring regions are too small, follow-up (or secondary) migrants will (re)locate to places where they can find a large network (again, mostly the capital), nullifying the policy's objective. Rather, such policies should distribute migrants from the same sending country into clusters of regions, where the network in the neighbouring regions of the cluster increases the attractiveness of the initial placement location.

The results in this paper also point to a strong 'lock-in effect': the current spatial distribution of migrants in part determines the future regional patterns of migration. This implies that migration laws affect not only present but also future migration flows. To provide a recent example, after the 2004 EU enlargement, EU-15 countries adopted heterogeneous restrictions on the movement of labour. Although these restrictions were allowed only during a transition period (and had to be suspended by 2011 at the latest), the heterogeneous use of these restrictions across member countries can be expected to have long-run effects on the patterns of migration from the new member states.

The relevance of these results go far beyond the topic addressed here: as migrant networks are characterized by a pronounced spatial structure, analysing the effects of migrants may lead to biased results and incorrect policy conclusions if the spatial structure of immigrant networks is not explicitly accounted for. This limitation may not only affect analyses of migrants' location decisions: related empirical studies on the effect of the regional stock of migrants on the national composition of firms' workforces (Joona & Wadensjö, 2009) or on firms' export behaviour (Pennerstorfer, 2016) are supposedly affected by a similar omitted-variables problem. The challenging task when analysing particular effects of immigrants is that the spatial scale (or the distance decay) of potential spillover effects may differ substantially. For instance, effects of immigration networks on innovation activities will appear only locally, as regular face-to-face contact is necessary to exchange ideas, while financial assistance for entrepreneurial activities may be provided over longer distances.

ACKNOWLEDGEMENTS

Congress of the European Regional Science Association, the 26th Congress of the European Economic Association, the 65th European Meeting of the Econometric Association, and the 2010 Winterseminar of the Gesellschaft für Regional forschung; as well as seminar participants at the University of Innsbruck, the University of Salzburg and the Vienna University of Economics and Business for helpful comments on an earlier version of this paper.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

SUPPLEMENT DATA

Supplemental data for this article can be accessed at https://doi.org/10.1080/00343404.2017.1380305.

NOTES

1. We thank an anonymous referee for pointing this out. The importance of expectations in migration decisions is empirically investigated by Bertoli, Brücker, and Fernández-Huertas Moraga (2016).
2. Some notable exceptions are: Pedersen et al. (2008), who estimated the determinants for migration flows to 22 Organisation for Economic Co-operation and Development (OECD) countries; Ruyssen and Rayp (2014), who investigated migration flows between 43 sub-Saharan African countries; Geis, Uebelmesser, and Werding (2013), who analysed the migrants' choice between four OECD countries (France, Germany, the UK and the United States); as well as the single-country studies by Damm (2009), Åslund (2005) and Jayet et al. (2016).
3. Imagine, for example, a sending country with two different ethnic groups, A and B, where a civil war has driven many members of ethnic group A abroad. With fixed coefficients, the error terms would then be positively correlated across regions with large migrant/refugee networks for members of ethnic group A, while they would not be correlated (or even negatively correlated) across the same regions for members of ethnic group B.
4. The application of the RPL is not limited to cases where individual data are available: just as an individual-level CL model can be estimated using Poisson regression, an aggregate-level Poisson model can be estimated using CL after appropriate changes in the data structure. Since CL is a special case of RPL (i.e., an RPL with fixed parameters only), this also applies to the RPL. Thus, the RPL model could also be applied to aggregate data to mitigate the problem of multilateral resistance. The equivalence between aggregate- and individual-level models (and the specification used here) rests on the assumption that the effects of regional-specific variables are not affected by the level of individual characteristics. Estimations on sub-groups defined by education, age and gender (see section B.5 and Table B7 in Appendix B in the supplemental data online) show that the results hardly differ between individuals with different characteristics, supporting this assumption.

5. Overseas territories, the relatively remote Canary Islands and the Azores and Madeira island regions as well as the Spanish exclaves Ceuta and Melilla are not considered here. Åland as well as the UK's 'Highlands and Islands' and 'North Eastern Scotland' regions must be excluded because of lack of data. Denmark must be considered a single NUTS-2 region.
6. The total number of observations used in the conditional and random parameters logit models is 2,659,140 (= 16,830 × 158) because these models require one observation per alternative for each individual.
7. Unfortunately, a more detailed differentiation by year of arrival in the host country is not available in the EU-LFS microdata at the authors' disposal.
8. Therefore, we follow a suggestion by Gibbons and Overman (2012), who argue that a spatial autoregressive (SAR) model (including spatial lags of the endogenous variable) should be abandoned in favour of SLX models in many situations due to identification problems. We do not include spatially lagged networks of recent (i.e., 1998–2007) migrants (cf. Ruyssen & Rayp, 2014), as it will take some time for newly arrived migrants to provide things such as ethnic goods or information externalities to other members of the network.
9. However, for example, Patacchini and Zenou (2012) used networks in neighbouring regions as a proxy for weak ties when estimating the effect of networks on the employment probability. A similar approach is applied by Jayet et al. (2016), who focused on directly adjacent locations only.
10. The effects of characteristics of the sending countries cannot be estimated because they do not vary over alternatives.
11. A variable indicating a common border between sending and receiving countries is not included in the regression because only a small number of sending countries share a common border with the EU-13. If included, a common border dummy is only significant at the 10% level, while all other coefficients remain virtually unchanged. For a more detailed discussion, see section B.6 in Appendix B in the supplemental data online.
12. For the results of a regression where the effects of common language, colonial ties and distance are allowed to vary over decision-makers, see section B.2 in Appendix B in the supplemental data online. The standard deviations of their estimated random parameters are not significantly different from zero, which implies that the parameters of these control variables are indeed fixed and not random. In addition, section B.2 also contains the results of a regression that includes random receiving country effects in addition to regional fixed effects. However, the results in this paper are highly robust to this change in specification.
13. Although it would also be possible to estimate the model separately by country of origin, the regressions in this paper were estimated on a pooled sample of migrants because the main interest is on the average effect of the

network variables for migrants from all origin countries. Furthermore, a pooled sample was chosen to maintain comparability with cross-sectional gravity models of migration that include migrants from several sending countries. For additional estimations (e.g., including migrant networks based on common language or on subsets defined by education, age and gender) that support the robustness of the results, see Appendix B in the supplemental data online.

14.   Former colonial ties are concentrated in a small number of receiving countries (mainly the UK, France and, to a lesser extent, Portugal and the Netherlands). Migrants from Commonwealth countries were allowed to migrate freely to the UK until 1962. Up to 1973, immigration was easier for citizens of these countries than for those from other non-European countries (e.g., Hansen, 2000, p. 316). France, on the other hand, signed bilateral agreements in the 1960s with a number of countries (including former colonies such as Algeria, Morocco and Tunisia) to facilitate (labour) immigration. Therefore, France offered those countries preferential terms until the suspension of these treaties in 1974 (e.g., de Lary, 2004).

15.   Ortega and Peri (2009) provide evidence that more restrictive entry laws reduce immigration flows.

16.   The estimated parameters in a logit model are $\beta = \beta^*/\sigma$, where β^* is the 'true' parameter of the utility function; and σ^2 is the variance of the unobserved portion of utility expressed as a multiple of $\pi^2/6$ (the 'scale parameter'). If more explanatory variables enter the model, the unobserved portion of utility decreases, leading to a decline in σ. But as σ decreases, β can be expected to increase (as long as the newly included explanatory variables are uncorrelated with the previously excluded regressors) (Train, 2009, pp. 40f.).

17.   The correlations between the (log) networks in s, N_s^1, N_s^2 and N_s^C range from 0.432 to 0.716. The correlations between $\text{Network}_{js}(s)$ and the networks in $N_s^{1'}$ and $N_s^{2'}$ are 0.160 and 0.206 respectively.

ORCID

Klaus Nowotny http://orcid.org/0000-0003-0488-0570
Dieter Pennerstorfer http://orcid.org/0000-0002-6191-4397

REFERENCES

Åslund, O. (2005). Now and forever? Initial and subsequent location choices of immigrants. *Regional Science and Urban Economics*, *35*(2), 141–165. doi:10.1016/j.regsciurbeco.2004.02.001

Bartel, A. P. (1989). Where do the new U.S. immigrants live? *Journal of Labor Economics*, *7*(4), 371–391. doi:10.1086/298213

Bauer, T., Epstein, G. S., & Gang, I. N. (2000). *What are migration networks?* (IZA Discussion Paper No. 200). Bonn: Institute for the Study of Labour (IZA).

Bauer, T., Epstein, G. S., & Gang, I. N. (2005). Enclaves, language and the location choice of migrants. *Journal of Population Economics*, *18*(4), 649–662. doi:10.1007/s00148-005-0009-z

Bauer, T., Epstein, G. S., & Gang, I. N. (2007). The influence of stocks and flows on migrants' location choices. *Research in Labor Economics*, *26*, 199–229. doi:10.1016/S0147-9121(06)26006-0

Beine, M., Bertoli, S., & Fernández-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. *World Economy*, *39*(4), 496–512. doi:10.1111/twec.12265

Beine, M., Docquier, F., & Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, *95*(1), 30–41. doi:10.1016/j.jdeveco.2009.11.004

Beine, M., Docquier, F., & Özden, Ç. (2015). Dissecting network externalities in international migration. *Journal of Demographic Economics*, *81*(4), 379–408. doi:10.1017/dem.2015.13

Bertoli, S., Brücker, H., & Fernández-Huertas Moraga, J. (2016). The European crisis and migration to Germany. *Regional Science and Urban Economics*, *60*, 61–72. doi:10.1016/j.regsciurbeco.2016.06.012

Bertoli, S., & Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, *102*, 79–100. doi:10.1016/j.jdeveco.2012.12.001

Bertoli, S., & Fernández-Huertas Moraga, J. (2015). The size of the cliff at the border. *Regional Science and Urban Economics*, *51*, 1–6. doi:10.1016/j.regsciurbeco.2014.12.002

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

Chiswick, B., & Miller, P. W. (2005). Do enclaves matter in immigrant adjustment? *City & Community*, *4*(1), 5–35. doi:10.1111/j.1535-6841.2005.00101.x

Christiadi, C., & Cushing, B. (2008). The joint choice of an individual's occupation and destination. *Journal of Regional Science*, *48*(5), 893–919. doi:10.1111/j.1467-9787.2008.00570.x

Dahlberg, M., & Eklöf, M. (2003). *Relaxing the IIA assumption in locational choice models: A comparison between conditional logit, mixed logit, and multinomial probit models* (Working Paper No. 2003:9). Uppsala: Department of Economics, Uppsala University.

Damm, A. P. (2009). Determinants of recent immigrants' location choices: Quasi-experimental evidence. *Journal of Population Economics*, *22*(1), 145–174. doi:10.1007/s00148-007-0148-5

Davies, P. S., Greenwood, M. J., & Li, H. (2001). A conditional logit approach to U.S. state-to-state migration. *Journal of Regional Science*, *41*(2), 337–360. doi:10.1111/0022-4146.00220

de Lary, H. (2004). Bilateral labour agreements concluded by France. In *Migration for employment* (pp. 43–54). Paris: OECD Publ.

EUROSTAT. (2009). *Labour force survey in the EU, candidate and EFTA countries. Main characteristics of the national surveys* (2009th ed.). Luxembourg: Office for Official Publications of the European Communities.

EUROSTAT. (2013). *EU labour force survey – Statistics explained*. Retrieved November 25, 2013, from http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_labour_force_survey

Geis, W., Uebelmesser, S., & Werding, M. (2013). How do migrants choose their destination country? An analysis of institutional determinants. *Review of International Economics*, *21*(5), 825–840. doi:10.1111/roie.12073

Gibbons, S., & Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, *52*(2), 172–191. doi:10.1111/j.1467-9787.2012.00760.x

Gottlieb, P. D., & Joseph, G. (2006). College-to-work migration of technology graduates and holders of doctorates within the United States. *Journal of Regional Science*, *46*(4), 627–659. doi:10.1111/j.1467-9787.2006.00471.x

Grogger, J., & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, *95*(1), 42–57. doi:10.1016/j.jdeveco.2010.06.003

Gross, D. M., & Schmitt, N. (2003). The role of cultural clustering in attracting new immigrants. *Journal of Regional Science*, 43(2), 295–318. doi:10.1111/1467-9787.00300

Guimarães, P., Figueirdo, O., & Woodward, D. (2003). A tractable approach to the firm location decision problem. *Review of Economics and Statistics*, 85(1), 201–204. doi:10.1162/003465303762687811

Guimarães, P., Figueirdo, O., & Woodward, D. (2004). Industrial location modeling: Extending the random utility framework. *Journal of Regional Science*, 44, 1–20. doi:10.1111/j.1085-9489.2004.00325.x

Hansen, R. (2000). *Citizenship and immigration in postwar Britain: The institutional origins of a multicultural nation*. New York: Oxford University Press.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: The state of practice. *Transportation*, 30(2), 133–176. doi:10.1023/A:1022558715350

Jaeger, D. A. (2007). Green Cards and the location choices of immigrants in the United States. *Research in Labor Economics*, 27, 131–183. doi:10.1016/S0147-9121(07)00004-0

Jayet, H., Rayp, G., Ruyssen, I., & Ukrayinchuk, N. (2016). Immigrants' location choice in Belgium. *Annals of Regional Science*, 57(1), 63–89. doi:10.1007/s00168-016-0761-x

Joona, P. A., & Wadensjö, E. (2009). The employees of native and immigrant self-employed. In A. F. Constant, K. Tatsiramos, & K. F. Zimmermann (Eds.), *Ethnicity and labor market outcomes* (Research in Labor Economics Vol. 29) (pp. 229–250). Bingley: Emerald Group.

LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: Chapman & Hall/CRC.

Marschak, J. (1960). Binary choice constraints on random utility indications. In K. Arrow (Ed.), *Stanford symposium on mathematical methods in the social sciences* (pp. 312–322). Stanford: Stanford University Press.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegriono, A., & Taylor, J. (1993). Theories of international migration: A review and appraisal. *Population and Development Review*, 19(3), 431–466. doi:10.2307/2938462

Mayer, T., & Zignano, S. (2011). *Notes on CEPII's distances measures: The GeoDist database* (CEPII Working Paper No. 2011-25). Paris: Centre d'Etudes Prospectives et d'Informations Internationales (CEPII).

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

McFadden, D., & Train, K. E. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470. doi:10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the U.S. labor market. *Quarterly Journal of Economics*, 118(2), 549–599. doi:10.1162/0033553033321675455

Ortega, F., & Peri, G. (2009). *The causes and effects of international migrations: Evidence from OECD countries 1980*–2005 (NBER Working Paper No. 14833). Cambridge, MA: National Bureau of Economic Research (NBER).

Ortega, F., & Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies*, 1(1), 47–74. doi:10.1093/migration/mns004

Patacchini, E., & Zenou, Y. (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics*, 42(6), 938–949. doi:10.1016/j.regsciurbeco.2012.01.004

Pedersen, P. J., Pytlikova, M., & Smith, N. (2008). Selection and network effects – Migration flows into OECD countries 1990–2000. *European Economic Review*, 52(7), 1160–1186. doi:10.1016/j.euroecorev.2007.12.002

Pennerstorfer, D. (2016). Export, migration and costs of trade: Evidence from Central European firms. *Regional Studies*, 50(5), 848–863. doi:10.1080/00343404.2014.947565

Ruyssen, I., & Rayp, G. (2014). Determinants of intraregional migration in sub-Saharan Africa 1980–2000. *Journal of Development Studies*, 50(3), 426–443. doi:10.1080/00220388.2013.866218

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4), 641–658. doi:10.1162/rest.88.4.641

Schmidheiny, K., & Brülhart, M. (2011). On the equivalence of location choice models: Conditional logit, nested logit and Poisson. *Journal of Urban Economics*, 69(2), 214–222. doi:10.1016/j.jue.2010.09.004

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). New York: Cambridge University Press.

Zavodny, M. (1999). Determinants of recent immigrants' locational choice. *International Migration Review*, 33(4), 1014–1030. doi:10.2307/2547361

Zorlu, A., & Mulder, C. H. (2008). Initial and subsequent location choices of immigrants to the Netherlands. *Regional Studies*, 42(2), 245–264. doi:10.1080/00343400601145210