EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

University of Tübingen
Working Papers in
Economics and Finance

No. 46

Co-national and Transnational Networks in International Migration to Spain

by

Nina Neubecker & Marcel Smolka

Faculty of Economics and Social Sciences
www.wiwi.uni-tuebingen.de

Co-national and Transnational Networks in International Migration to Spain*

Nina Neubecker[a] Marcel Smolka[b]
University of Tübingen University of Tübingen

October 2012

Abstract

This paper provides evidence that transnational networks, defined as networks operating across nationalities, are shaping observed patterns of international migration. In a stylized model of migration with random friendship formation, individuals from a given origin country are attracted to destinations hosting large migrant communities from countries which are culturally and geographically close to their own origin country. In addition, the attracting force of a large community of co-national migrants is the larger, the larger the community of migrants from other culturally proximate countries in the same destination. Both predictions are supported by aggregate migration data on international migration to Spain, detailed by origin country and destination province. Our findings imply that the literature estimating network effects in migration has been overly restrictive in its definition of migrant networks.

Keywords: international migration, co-national and transnational friendships, network effect, Spain.
JEL classification: F22

[a]**Corresponding author:** Faculty of Economics and Social Sciences, University of Tübingen, Mohlstraße 36, 72074 Tübingen, Germany. E-mail: nina.neubecker@uni-tuebingen.de.
[b]Faculty of Economics and Social Sciences, University of Tübingen, Nauklerstraße 47, 72074 Tübingen, Germany. E-mail: marcel.smolka@uni-tuebingen.de.

1 Introduction

A large and growing body of literature has shown that informal social networks alleviate the burden of migration in that they serve as important transmission channels for information on jobs and housing at destination. The increased economic globalization observed over the past decades, however, is changing the network structure of friends and contacts, leading to more frequent interactions among individuals with different nationalities.[1] As modern technologies allowing for cheap communication reach ever more population groups also in less advanced countries, transnational friendships can easily be cultivated or quickly revived.[2]

In this paper, we provide evidence that transnational networks, i.e., networks operating *across nationalities*, are shaping observed patterns of international migration. To guide our empirical analysis, we set up a stylized model of migration which formalizes, in the simplest possible way, the idea of friendship formation both between individuals who share the same nationality and between those who do not. We test novel empirical predictions of how migrant networks impact on the scale of migration, using rich migration data from the Spanish Instituto Nacional de Estadística (INE) on the recent immigration boom to Spain.

Our stylized model distinguishes between what we call *pioneer migration* to Spain in the first period and *follow-up migration* to Spain in the second period. Before pioneer migration takes place, friendships among individuals are formed randomly, depending on the cultural and geographical proximity of their origin countries.[3] Individuals having friends among the pioneer migrants in Spain are

[1]The world has in fact reached unprecedented levels of globalization. International trade flows and global FDI stocks are both estimated to equal 30% of world output in 2010; see UNCTADstat at `http://unctadstat.unctad.org/`, accessed on 09/03/2011. In the same year, the number of foreign-born individuals was 3.1% of the total world population (United Nations, 2009).

[2]By the term "transnational friendship" we mean any friendship involving more than one nationality. For the sake of simplicity, we refer to all types of social ties as friendships in the following.

[3]In culturally and geographically proximate countries, the formation of transnational friendships may be more likely than perhaps expected. For example, in the year 2000, 7.0% of all individuals living in Costa Rica held a foreign nationality of another country in Latin America or the Caribbean. The same number for larger countries such as Venezuela (3.2%), Paraguay (3.0%), or Argentina (2.9%) is smaller but still not negligible; see the World Bank's Global

more likely to become follow-up migrants at the end of the second period, given that they have access to private information on local conditions they would otherwise not have.[4] This implies that individuals from a given origin country are attracted to destinations hosting large migrant communities from countries which are culturally and geographically close to their own origin country. This prediction is supported by our data.

Pioneer migrants, as opposed to non-migrants, form new friendships in their destination location at the beginning of the second period (before follow-up migration). In expectations, more friendships are formed among culturally similar than among culturally dissimilar pioneer migrants. Through second-order friends (friends of friends), an enhanced set of friendship relations among the pioneer migrants in a given destination makes an individual contact in this destination more valuable from the perspective of those left behind. This implies that the attracting force of a large community of co-national migrants should be the larger, the larger the community of migrants from other culturally close countries in the same destination. The evidence presented in this paper supports this positive interaction.

Much of the theoretical literature on network effects in migration *assumes* that pioneer migrants reduce the migration costs for those left behind, both through monetary and non-monetary assistance to follow-up migrants (Carrington et al., 1996; Chau, 1997). By contrast, Stark & Wang (2002) investigate the *incentives* for pioneer migrants to provide this assistance to their followers, finding that high-skilled pioneer migrants optimally choose to subsidize low-skilled follow-up migrants due to a beneficial skill separation effect.[5] In our aggregate data, we do not observe whether and how

Bilateral Migration Database at `http://data.worldbank.org/data-catalog/global-bilateral-migration-database`, accessed on 26/09/2012.

[4] Evidence from the Spanish National Immigrant Survey 2007 (NIS) supports the importance of having contacts at destination in the decision to migrate to Spain. For the year 2007, about 80% of all surveyed immigrants in Spain report that they had a contact person in Spain upon their arrival. For 70% of those who had somebody to turn to in Spain, the contact persons were family members. Yet, friends (30%), acquaintances (8%), business persons or legal agents (3%), intermediaries or non-official persons (1%), and other persons (2%) also played a role.

[5] In earlier work, Stark (1999) shows that high-skilled migrants may on the contrary want to prevent follow-up migration of low-skilled workers if employers at destination do not know migrants' skill levels.

friendships can actually help reduce the cost of migration, nor do we observe immigrants' actual friendships, whether or not they bridge two nationalities. Instead, we take a probabilistic approach towards friendship formation which can help rationalize correlations observed in aggregate migration data. The idea that transnational networks impact on migrants' location choices is well in line with descriptive evidence on the geographical distribution of immigrants in Spain. In particular, we see that immigrants in Spain originating from the same world region tend to cluster in specific Spanish provinces. Furthermore, we show that, in general, the geographical settlement pattern of immigrants from two different countries is more similar, the smaller the geographical distance between the two countries considered.

Our paper is most closely related to the literature concerned with estimating network effects in migration. This literature has been developed against the backdrop of a large body of sociological studies highlighting the importance of informal social networks for many types of human activities including migration. Massey et al. (1993, 448) define migrant networks as "[...] sets of interpersonal ties that connect migrants, former migrants, and nonmigrants in origin and destination areas through ties of kinship, friendship, and shared community origin". This definition is not incompatible with the idea of transnational networks invoked in our paper, given that such networks may well be rooted in kinship, friendship, and some temporarily shared community origin. By focusing on common origin defined at the country or sub-country level, however, all empirical studies we are aware of have ignored the role of transnational networks in shaping aggregate patterns of migration.[6] For instance, Bauer et al. (2007, 2009) use micro-level data on Mexican immigrants in the U.S. and construct measures for migrant networks based on common village origin. Studies using macro-level data define migrant networks in terms of a common country of origin, country of birth, or nationality (Hatton, 1995;

[6] Controlling for the migrant network of co-national individuals, Åslund (2005) finds that immigrants in Sweden are attracted to regions with a large number of foreigners relative to the total population. However, he does not distinguish among different nationalities of these foreigners. In Neubecker et al. (2012), we intend to control for transnational network effects through fixed effects, but we do not identify them.

Clark et al., 2007; Lewer & Van den Berg, 2008; Pedersen et al., 2008; Beine et al., 2011; Neubecker et al., 2012). They find strong support for the importance of networks in determining the scale of migration.[7] Few empirical studies look at the effect of migrant networks defined at the family level, exploiting more detailed information on the precise type of network ties. Davis et al. (2002) find that closer kinship bonds result in a larger impact of the migrant network. Dolfin & Genicot (2010) argue that family networks provide information on jobs and act as a source of credit, while community networks are more important sources of information on the border-crossing as such.

The remainder of this paper is organized as follows. In section 2 we provide a descriptive look at clustering and settlement patterns of immigrants in Spain in relation to the cultural and geographical proximity of their origin countries. In section 3 we present a stylized model of migration with random friendship formation that guides us in our subsequent empirical analysis of the role of transnational networks in international migration to Spain. Section 4 describes our empirical analysis and discusses the estimation results. Section 5 concludes.

2 Geographical Distribution Patterns of Migrants in Spain

This section provides some descriptive evidence on the geographical distribution of different migrant populations in Spain. This evidence suggests that migrants prefer to settle in provinces with large populations of migrants from countries that are located close to their own origin country. By migrants we mean people who live in Spain and hold a foreign nationality. Information on migrant stocks come from the Spanish Municipal Register and are freely available from the INE website.[8]

A first observation is that migrants, in their entirety, are not evenly distributed across Spanish

[7]In our own earlier work (Neubecker et al., 2012), we also find that migrant networks in Spain bias the skill structure of new immigrants towards the low-skilled individuals; see also Beine et al. (2011) and Beine et al. (2012) for immigration to the OECD and the U.S., respectively.

[8]For detailed information on all data sources used in this paper, see table A.1 in the appendix.

provinces. The four major destination provinces account for as much as 47% of all migrants registered at Spanish municipalities in the year 2009. These provinces are Madrid (18.8%), Barcelona (14.2%), Alicante (8.2%), and Valencia (5.6%) and rank also among the most populous provinces in general.

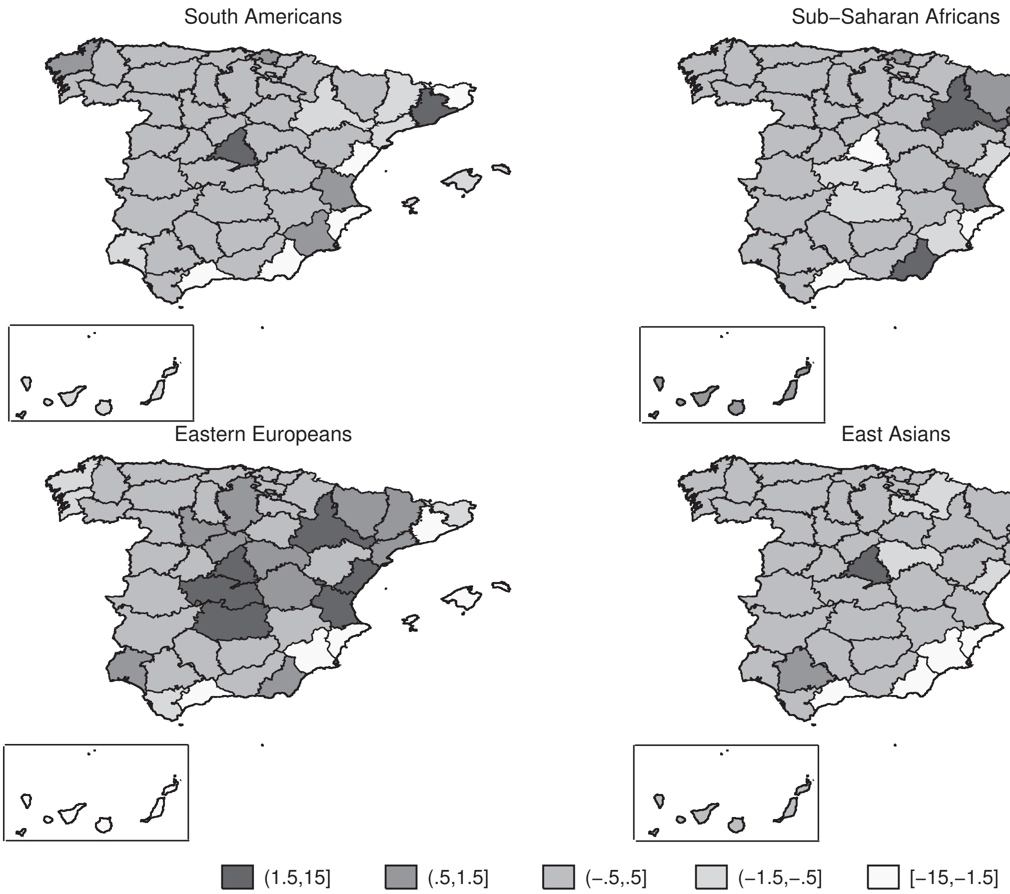
Another observation is that migrants originating from the same world region tend to concentrate in specific Spanish provinces. For instance, migrants from South America, Sub-Saharan Africa, Eastern Europe as well as East Asia are all significantly more concentrated in Madrid and Barcelona than Spanish nationals.[9] For each of the four world regions, the share of migrants residing in either of the two provinces exceeds the corresponding share of Spanish nationals by a number larger than 0.15. Migrants from these world regions also reside more often than Spanish nationals in certain Northern provinces (Vizcaya, Zaragoza, Girona) as well as in several provinces along the Spanish Mediterranean coast (Tarragona, Valencia, Alicante, Murcia, Málaga). We refer to this strong geographical concentration of migrants relative to Spanish nationals as *clustering*.

Apart from network effects leading to migration flows being path-dependent, we have no specific reason to believe that migrants exhibit relevant differences in the *degree* of clustering across world regions. In order to find out about differences in the settlement patterns across migrant groups, figure 1 compares the geographical distribution of migrants from each of the four world regions to the distribution of all migrants in Spain in the year 2009 (each time excluding migrants from the world region under consideration). Dark colors indicate a strong concentration of migrants from a given world region relative to all other migrants, while light colors indicate a relatively weak concentration. We see, for instance, that migrants from South America, Eastern Europe, and East Asia are more heavily clustered in Madrid than migrants from other world regions. The opposite holds true for migrants from Sub-Saharan Africa. In Barcelona, migrants from South America, Sub-Saharan

[9] South America is the most important origin region of migrants in Spain (1.6 million migrants in the year 2009). Eastern Europe ranks second (1.3 million migrants), Sub-Saharan Africa fifth (227,000 migrants), and East Asia sixth (155,000 migrants).

Africa, and East Asia are more strongly clustered than other migrants, while migrants from Eastern Europe are clustered less than migrants from other world regions. Further differences in the degree of concentration can be spotted for other provinces.

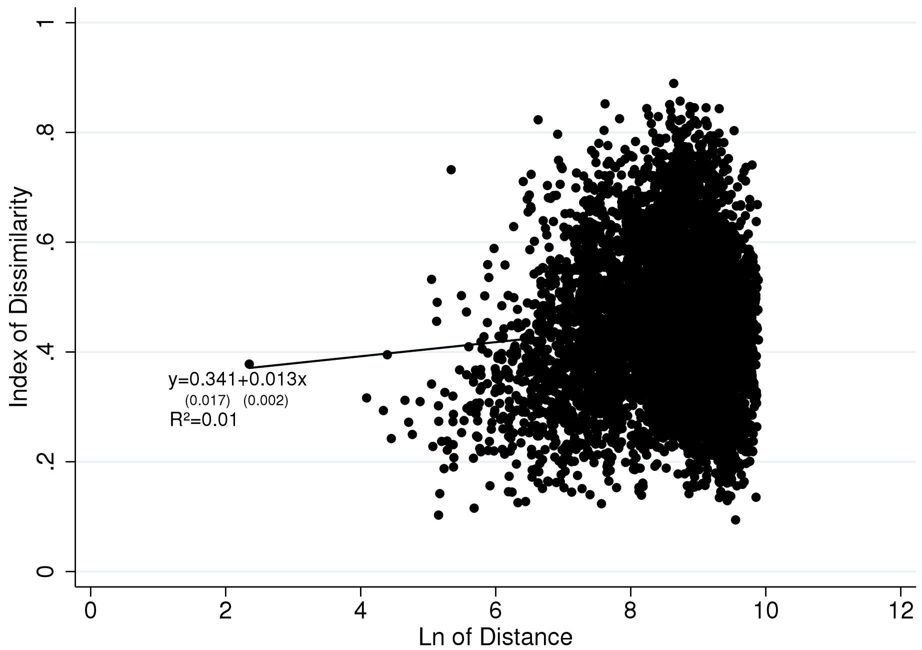
Figure 1: Differences in the Geographical Concentration of Migrant Populations in Spain, 2009†



† This figure illustrates differences in the geographical distributions of migrants in Spain from four different world regions relative to the distribution of all migrants in Spain in the year 2009 (each time excluding migrants from the world region under consideration). Dark colors indicate a strong concentration of migrants from a given world region relative to all other migrants, while light colors indicate a relatively weak concentration. The provinces Las Palmas and Santa Cruz de Tenerife are grouped together as Islas Canarias. *Source:* Authors' tabulations using data from INE.

We next take a slightly more formal approach to look at the relationship between the settlement patterns of migrants in Spain and the cultural and geographical proximity of their origin countries. In particular, we ask whether differences in the geographical distribution of migrants originating from any two countries correlate with the bilateral distance between these two countries. For this purpose, figure 2 plots the country-pair-specific index of dissimilarity à la Duncan & Duncan (1955) for any two migrant populations settled in Spain in the year 2009 against the log of the bilateral distance (measured in kilometers) between the two origin countries considered. The index of dissimilarity is a summary statistic for the differences in the geographical distributions of two populations. It is defined as $D = 0.5 \sum_1^k |x_i - y_i|$ where, in our case, x_i is the share of a certain migrant group residing in province i, y_i is the corresponding share of a second migrant group, and k is the total number of provinces in Spain. The index gives the share of migrants from the x-group who would have to move to other Spanish provinces in order to replicate the geographical distribution of migrants from the y-group; see Duncan & Duncan (1955, 211). Thus, D can only take on values in the unit interval, with higher numbers indicating stronger dissimilarity in location choices between the two migrant groups.

Figure 2: Index of Dissimilarity of Migrant Populations in Spain and Bilateral Distance between Origin Countries, 2009†



† This figure plots the country-pair-specific index of dissimilarity à la Duncan & Duncan (1955) for any two migrant populations settled in Spain in the year 2009 against the log of the bilateral distance (measured in kilometers) between the two origin countries considered. Larger values of the index of dissimilarity indicate stronger dissimilarity in the geographical distributions of two migrant populations. *Source:* Authors' tabulations using data from INE and CEPII.

The linear best fit in figure 2 indicates a positive albeit small correlation between the dissimilarity index and the bilateral distance variable (statistically significant at the 1% level), showing that migrants from a given origin country tend to settle in provinces where other migrants from geographically proximate countries settle as well. This positive correlation is in line with our earlier observation that migrants from a given world region are more heavily clustered in certain provinces than all other migrants.

3 A Theory of Migration with Random Friendship Formation

In this section, we set up a stylized model of migration which guides us in our empirical analysis of transnational networks and their effects on international migration to Spain. We assume that a random process leads to the formation of friendships both between individuals who share the same nationality and those who do not. We refer to the former as *co-national friends* and to the latter as *transnational friends*. Friendships are formed at the beginning of period 1, with individuals migrating to Spain at the end of this period. This *pioneer migration* is given exogenously and independent of the number of friends individuals have acquired before. At the beginning of period 2, new random friendships are formed in Spain, with further individuals migrating to Spain at the end of this period. This *follow-up migration* is influenced by the network structure of friends developed in both periods. It is precisely this migration that we seek to explain in our empirical analysis.

Friendship Formation and Pioneer Migration in Period 1

Individuals differ in terms of their nationalities, indexed by i or $j = 1, \ldots, I$. We denote the number of individuals with nationality i by N_i. Let a friendship between any i-individual and any j-individual be formed randomly with the following probability:

$$p_{ij} = \frac{\eta_{ij}}{N_w}, \tag{1}$$

where $\eta_{ij} > 0$ measures the cultural and geographical proximity of the origin countries of individuals with nationalities i and j, and $N_w = \sum N_i$ is the world population. Friendship formation between any two individuals with nationalities i and j is more likely, the larger the cultural and geographical proximity of their origin countries; it is less likely, the larger the world population. For $\eta_{ij} = \bar{\eta}$, an individual's expected total number of friends is equal to $\bar{\eta}$, irrespective of the size of the world population.

Final migration destinations are indexed by $k = 1, \ldots, K$.[10] At the end of the first period some individuals of each nationality migrate to destination k. We denote the number of j-individuals living in destination k at the beginning of period 2 by M_{jk} and assume that this pioneer migration is independent of the number of friends an individual has acquired before. Consider a randomly drawn individual o with nationality i. Given the binomial nature of the friendship formation process, the probability that individual o has exactly x friends with nationality j at destination k is equal to:

$$\Pr(x^o_{jk} = x) = \binom{M_{jk}}{x} p^x_{ij} \left(1 - p_{ij}\right)^{M_{jk}-x}, \tag{2}$$

where x^o_{jk} is individual o's number of friends with nationality j in destination k, and M_{jk} is the number of pioneer migrants with nationality j who are settled in destination k. It is well known that for large M_{jk} and small p_{ij} the binomial distribution is approximately equal to the Poisson distribution. We can thus write:

$$g(x; \lambda_{ijk}) \equiv \Pr(x^o_{jk} = x) \approx \frac{\exp(-\lambda_{ijk})\lambda^x_{ijk}}{x!}, \tag{3}$$

[10]For expositional reasons, we do not explicitly include each migrant's origin country into the set of migration destinations, even though some individuals in each country do not migrate abroad. However, it is possible to modify the model setup so as to include each individual's origin country into the set of migration destinations.

where $\lambda_{ijk} \equiv M_{jk}p_{ij} = M_{jk}\eta_{ij}/N_w$ is equal to the expected number of j-friends in destination k of a randomly drawn i-individual. This number is increasing in the number of pioneer migrants with nationality j settled in destination k, but decreasing in the total number of individuals, irrespective of their nationalities. The variable λ_{ijk} represents the community of j-migrants in destination k, scaled by the world population and weighted by the cultural and geographical proximity of the origin countries of individuals with nationalities i and j before pioneer migration has taken place. By slightly abusing terminology, we will refer to λ_{ijk} as the ex-ante network of j-migrants in k, from the viewpoint of individuals with nationality i.

Lemma 1. *An i-individual's probability of having at least $\bar{x}$ j-friends in destination k is increasing in the ex-ante network of j-migrants in k, λ_{ijk}.*

Proof. In order to prove that $\partial \Pr(x_{jk}^o \geq \bar{x})/\partial\lambda_{ijk}$ is positive for any $\bar{x}$ and any individual o with nationality i, it is sufficient to show that $g(x;\lambda)$ exhibits the monotone likelihood ratio property according to which any two probability mass functions $g(x;\lambda)$ and $g(x;\lambda')$ with $\lambda > \lambda'$ satisfy

$$\frac{g(x_1;\lambda)}{g(x_1;\lambda')} \geq \frac{g(x_0;\lambda)}{g(x_0;\lambda')} \tag{4}$$

for any $x_1 > x_0$. We compute

$$\frac{g(x;\lambda)}{g(x;\lambda')} = \exp(\lambda' - \lambda)\left(\frac{\lambda}{\lambda'}\right)^x \tag{5}$$

and notice that this expression is strictly increasing in x since $\lambda > \lambda'$. □

Friendship Formation in Period 2

Individuals settling in a new location radically change their social environment and thus interact with a set of individuals different from the one in their previous location. We therefore assume that pioneer migrants, as opposed to non-migrants, form new friendships in their destination location, while preserving all their friendships formed in the first period. A friendship between two pioneer migrants with nationalities i and j living in destination k is formed with the following probability:

$$p'_{ijk} = \frac{\eta'_{ij}}{M_k}, \tag{6}$$

where η'_{ij} is the cultural proximity of pioneer migrants with nationalities i and j, and $M_k \equiv \sum_i M_{ik}$ is the size of the total population living in k at the beginning of period 2. Hence, as in period 1, friendship formation in period 2 follows a binomial distribution. The probability that a pioneer migrant q holding nationality i and living in destination k forms exactly x new friendships with j-individuals in k is approximately given by:

$$h(x; \kappa_{ijk}) \equiv \Pr(x^q_{jk} = x) \approx \frac{\exp(-\kappa_{ijk})\kappa^x_{ijk}}{x!}, \tag{7}$$

where $\kappa_{ijk} \equiv M_{jk}p'_{ijk} = M_{jk}\eta'_{ij}/M_k$ is equal to individual q's expected number of new j-friends in destination k. We will refer to the variable κ_{ijk} as the ex-post network of j-migrants in k, from the viewpoint of individuals with nationality i living in destination k.

Lemma 2. *Consider a pioneer migrant in destination k with nationality i. Her probability of having at least $\bar{x}$ new j-friends in destination k is increasing in the ex-post network of j-migrants in k, κ_{ijk}.*

Proof. Similar to the proof of Lemma 1. □

Follow-up Migration in Period 2

At the end of period 2, a random individual o with nationality i who has not migrated in period 1 compares the utility U^o_k of all K destinations and moves to the destination from which she derives the highest utility (follow-up migration):

$$k^o = \text{argmax}(U_1^o, \ldots, U_K^o). \tag{8}$$

The probability that individual o moves to destination k is equal to the probability that this individual associates the largest utility with moving to destination k:

$$\Pr(k^o = k) = \Pr(U_k^o > U_{k'}^o \; \forall \; k' \neq k). \tag{9}$$

Informational frictions prevail in the labor markets of all destinations $k = 1, \ldots, K$. More specifically, information on job vacancies are not publicly available but are partly circulated among friends. Finding a job in some destination k is less difficult for individual o, if she has access to some of this information through her network of friends.

We assume that U_k^o is a continuous and differentiable function of the relevant information available to individual o, for simplicity assumed one-dimensional and measured by the variable T_k^o, with $\partial U_k^o / \partial T_k^o > 0$. By continuity and monotonicity, there must be a critical level $\tilde{T}_k^o$ for which $U_k^o > U_{k'}^o \; \forall \; k' \neq k$. Hence, we can rewrite equation (9) as:

$$\Pr(k^o = k) = \Pr(T_k^o \geq \tilde{T}_k^o). \tag{10}$$

Notice that the threshold value $\tilde{T}_k^o$ depends on all utility-relevant characteristics of all destinations.

We assume that T_k^o is a continuous and differentiable function of the number of individual o's *direct* and *indirect* j-friends in destination k. The number of individual o's *direct* j-friends in k is denoted by x_{jk}^o. It is a simple count of her direct friendship relations with j-individuals in k. The number of individual o's *indirect* j-friends in k is denoted by y_{jk}^o. It counts the direct j-friends in k

of individual o's direct friends in k.[11] We write:

$$T_k^o \equiv T(x_{1k}^o, \ldots, x_{Ik}^o, y_{1k}^o, \ldots, y_{Ik}^o) \tag{11}$$

and assume, plausibly, that $\partial T_k^o / \partial x_{jk}^o > 0$ as well as $\partial T_k^o / \partial y_{jk}^o > 0$ for all $j = 1, \ldots, I$.

Proposition 1. *The share of follow-up migrants $m_{ik}/(N_i - \sum_k M_{ik})$ with nationality i to some destination k is the larger, the larger the ex-ante network of j-migrants in k, λ_{ijk}. For a given non-empty set of direct friendships between i-individuals who have not migrated in the first period and some pioneer migrants (of whatever nationality) in k, the share is also increasing in the ex-post network of j-migrants in k.*

Proof. We show first that $\Pr(T_k^o \geq \tilde{T}_k^o)$ is increasing in λ_{ijk} for any individual o with nationality i. To see this, note that a larger ex-ante network of j-migrants in k increases the probability of having at least $\bar{x}$ direct j-friends in k (Lemma 1). Hence, the probability of having at least $\bar{x}$ indirect friends of all other nationalities in k increases as well, due to the friends of the direct j-friends in k. In addition, the probability of having at least $\bar{x}$ indirect j-friends in k increases even further due to Lemma 2. Aggregating over all individuals with nationality i who have not migrated in the first period proves the first part of Proposition 1, given that the set of pioneer migrants with nationality i is a random sample of all i-individuals. The second part of Proposition 1 follows from Lemma 2, because the probability of having at least $\bar{x}$ indirect j-friends in k is increasing in κ_{ijk} for any individual o with nationality i. □

[11] This definition implies that if individual o has no direct friends whatsoever in destination k, she cannot have any indirect friends in that same destination. However, through transnational friendships, she can have some indirect friends with nationality j in destination k even if she has no direct friends with nationality j in that same destination.

4 Empirical Analysis

In this section, we first describe our estimation strategy together with the data we use, and then present and discuss our estimation results. We conclude this section with a robustness analysis of the results obtained.

4.1 Estimation Strategy and Data

Setting up an estimable migration function in the spirit of Proposition 1 is not trivial. Some additional structure is needed here. As to the first part of Proposition 1, we distinguish between the ex-ante network of co-nationals, λ_{iik}, and that of all other nationalities, $\sum_{j\neq i}\lambda_{ijk}$. This allows us to discriminate between the network effect typically estimated in the literature and the network effect due to transnational friendships. As to the second part of Proposition 1, we focus exclusively on the interaction between the ex-ante network of co-nationals, λ_{iik}, and the sum of all ex-post migrant networks, $\sum_j \kappa_{ijk}$, although, strictly speaking, Proposition 1 purports a significant interaction between the networks of *any two* nationalities. Furthermore, we argue that the probability of friendship formation between pioneer migrants and native individuals at a given destination substantially differs from the probability of friendship formation between individuals of different migrant populations settled at the same destination. Therefore, we do not include the native population of a given migration destination into the sum of all ex-post migrant networks. We assume that the log of the share of follow-up migrants holding nationality i and moving from country i to destination k can be linearly approximated by the following expression:

$$\ln\left(\frac{m_{ik}}{N_i-\sum_k M_{ik}}\right)=\beta_0\ln\lambda_{iik}+\beta_1\ln\sum_{j\neq i}\lambda_{ijk}+\beta_2\ln\sum_j\kappa_{ijk}+\beta_3\ln\lambda_{iik}\times\ln\sum_j\kappa_{ijk}+\mu_{ik}, \quad (12)$$

where μ_{ik} captures all other bilateral as well as unilateral determinants of migration of individuals with nationality i to destination k. In the following we refer to the variable $\ln\lambda_{iik}$ as the *ex-ante* *co-national network*, to the variable $\ln \sum_{j \neq i} \lambda_{ijk}$ as the *ex-ante transnational network*, and to the variable $\ln \sum_j \kappa_{ijk}$ as the *ex-post transnational network*. In light of the first part of Proposition 1, we expect a positive effect of the ex-ante transnational network on migration, in addition to a (standard) positive effect of the ex-ante co-national network on migration. From the second part of Proposition 1, however, we expect this last effect to be increasing in the size of the ex-post transnational network. In turn, the ex-post transnational network itself should not exert an independent influence on migration if the ex-ante co-national network is very small (or zero).

Equation (12) can be rewritten as:

$$\begin{aligned} \ln m_{ik} &= \beta_0 \ln \left(\frac{\eta_{ii} M_{ik}}{N_w} \right) + \beta_1 \ln \left(\frac{1}{N_w} \sum_{j \neq i} \eta_{ij} M_{jk} \right) + \beta_2 \ln \left(\frac{1}{M_k} \sum_j \eta'_{ij} M_{jk} \right) \\ &\quad + \beta_3 \ln \left(\frac{\eta_{ii} M_{ik}}{N_w} \right) \times \ln \left(\frac{1}{M_k} \sum_j \eta'_{ij} M_{jk} \right) + \ln \left(N_i - \sum_k M_{ik} \right) + \mu_{ik}. \end{aligned} \tag{13}$$

In order to estimate equation (13), we draw on the same Spanish data as we do in Neubecker et al. (2012), complemented by data on the total world population and geographical distances, as well as by province-level data on GDP per capita, unemployment, and population density in Spain.[12] The sample considered comprises the 55 most important migrant-sending countries in terms of the aggregate number of migrants in Spain in the year 1996.[13]

The migration data come from the local registry information of Spanish municipalities provided through INE. A unique feature of these data is that they include both documented and undocumented immigrants. The reason for this is that through registration immigrants get access to free medical care under the same conditions as Spanish nationals, irrespective of whether or not they hold a valid residence permit; see Neubecker et al. (2012, 13) for details. Spain is divided into 52 provinces which

[12]The full internet sources of our data are listed in table A.1 in the appendix.

[13]These are listed in table A.2 in the appendix.

are nested in 19 regions. We exclude the enclaves Ceuta and Melilla due to their specific geographical location and thus end up with a total of 50 provinces considered in the estimations.[14]

The dependent variable is the log of the bilateral migration flow into Spanish provinces, obtained from the Spanish Residential Variation Statistics and aggregated from the beginning of 1997 until the end of 2006.[15] The size of the migrant population with nationality i at destination k, M_{ik}, is measured by the number of individuals with nationality i settled in k in the year 1996, as reported by the Spanish Municipal Register. We rely on population figures disaggregated by nationalities and Spanish provinces as of 1 May 1996. Since bilateral migrant stocks are zero in a non-negligible number of cases, we add one to these stocks when constructing the ex-ante co-national network variable.

The value of the world population, N_w, is observed for the year 1995 and taken from the United Nations Population Division. Furthermore, we proxy the cultural and geographical proximity between any two nationalities, η_{ij}, by the inverse of the geographical distance (in kilometers) between the most populous cities of the corresponding countries. We argue that cultural proximity, including linguistic proximity, is closely related to geographical proximity. Hence, we also measure the cultural proximity between pioneer migrants of different nationalities living in k, η'_{ij}, by the inverse of geographical distance. Furthermore, in order to account for the cultural and geographical proximity between individuals of the same nationality, η_{ii} and η'_{ii}, we rely on the distance between the respective country's two most populous cities. All these data come from the French Centre d'Études Prospectives et d'Informations Internationales (CEPII).

We control for several other potential determinants of the scale of bilateral migration to Spanish provinces, nested in the term μ_{ik}. Data on both trade and FDI are available from the Spanish

[14]See `http://www.ine.es/daco/daco42/codmun/cod_provincia.htm` (accessed on 04/17/2012) for a list of these provinces.

[15]We aggregate all migrants who registered at Spanish municipalities between 1 January 1997 and 31 December 2006 by their country of origin. Migrants are defined as individuals for whom the last country of residence (other than Spain) corresponds to their country of birth and nationality.

Ministry of Industry, Tourism and Trade. We measure bilateral trade flows by the sum of exports and imports (in Euros) between country i and Spanish province k in the year 1996. Data on FDI are observed as inflows into Spanish regions for the year 1997, detailed by country of origin. We add one to both variables before taking logs in order to keep observations with zero trade or FDI flows. We use three variables observed at the level of Spanish provinces in order to capture their attractiveness as migration destinations. First, we use (the log of) the provincial GDP per capita as a proxy for the average wage paid in each province. Second, we control for employment opportunities by including (the log of) the provincial employment rate, calculated as one minus the share of the unemployed in the total economically active population. Third, we proxy the degree of economic, cultural, and social activity in each province by its population density, measured as persons per square kilometer. All these data come from INE. Importantly, we employ these control variables rather than a set of province fixed effects in the estimations in order to separately identify the coefficients of the two transnational network variables, β_1 and β_2.[16]

We use the familiar fixed effects approach in estimating model (13), computing all variables as deviations from their country means (within-transformation). Since our migration data refer to a single destination country (Spain), this approach wipes out all determinants of migration with a unilateral dimension (referring to the sending country) or a bilateral dimension (involving the sending country and Spain). Among other things, we thereby control for the attractiveness of other migration destinations outside Spain.

More demanding specifications of our fixed effects model control for all effects specific to combinations of sending countries and Spanish regions. These are eliminated by computing all variables as deviations from their country-and-region means instead of country means. This approach greatly reduces the probability of omitted variables bias, because it controls for all determinants of bilat-

[16]Including province fixed effects would fix M_k in the variable $\ln \sum_j \kappa_{ijk}$ and thus render the two transnational network variables virtually collinear.

eral migration relevant for combinations of origin countries and destination regions in Spain; see also Neubecker et al. (2012). Given the potential endogeneity of the co-national network (Beine et al. 2011; Neubecker et al. 2012), we also present fixed effects estimations in which we instrument the ex-ante co-national network, $\ln \lambda_{iik}$, and its interaction with the ex-post transnational network, $\ln \lambda_{iik} \times \ln \sum_j \kappa_{ijk}$. We use the log of the number of people holding country i's nationality and emigrating from destination k in Spain to any other destination $l \neq k$ in Spain in the years 1988 and 1989, respectively, as excluded instruments.[17] Apart from the network effect itself, we cannot think of any channel through which the historical emigration flows within Spain can plausibly be expected to affect the recent immigration flows to Spain. Yet, we find that the historical emigration flows within Spain correlate significantly with the migrant communities in Spain even after a relatively long time.

4.2 Estimation Results

Table 1 shows the results from fixed effects (FE) estimations (columns a to d) and fixed effects two stage least squares (FE 2SLS) estimations (columns e to h). For each estimator, the first two columns (last two columns) control for country fixed effects (country-and-region fixed effects) through an adequate within-transformation. In these estimations, 5.7% of the observations need to be dropped due to zero migrant flows.[18] For the sake of comparison, we always report estimation results for a specification with all transnational network variables excluded, in addition to those for the full model specification as given by equation (13).

[17]Due to the interaction variable, the model has two endogenous variables. We therefore interact both excluded instruments with the ex-post transnational network and can thus conduct tests on overidentifying restrictions. As to the instrumental variables, we add one to the number of people before taking logs in order to keep observations with zero emigration flows.

[18]In the specifications controlling for country-and-region fixed effects, some additional observations are dropped because they refer to regions consisting of a single province in the dataset. The full matrix would consist of 55×50 observations.

Table 1: Estimation Results for the Full Model†

	Dependent Variable: Migration Inflow (Province-Level 1997-2006)							
	Fixed Effects				Fixed Effects Two Stage Least Squares			
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Ex-ante Co-national N.	0.786***	1.312***	0.606***	1.194***	0.962***	1.130***	0.836***	1.099***
(Province-Level 1996)	(0.024)	(0.112)	(0.028)	(0.098)	(0.057)	(0.196)	(0.074)	(0.129)
Ex-ante Transn. N.		0.570***		0.387***		0.392***		0.050
(Province-Level 1997)		(0.049)		(0.087)		(0.108)		(0.203)
Ex-post Transn. N.		1.496***		1.338***		0.691*		0.429
(Province-Level 1997)		(0.211)		(0.183)		(0.395)		(0.398)
Co-n. x *Ex-post Transn.*		0.054***		0.051***		0.026*		0.022*
(Province-Level 1997)		(0.009)		(0.008)		(0.015)		(0.013)
Trade Flow	0.027***	0.024***	0.004	0.003	0.023***	0.020***	0.005	0.004
(Province-Level 1996)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)
FDI Flow	0.022***	0.015***			0.012**	0.014***		
(Region-Level 1997)	(0.005)	(0.005)			(0.005)	(0.005)		
GDP per Capita	0.341**	0.740***	0.305	0.617***	0.397**	0.613***	0.381	0.344
(Province-Level 1997)	(0.171)	(0.179)	(0.224)	(0.232)	(0.173)	(0.194)	(0.236)	(0.280)
Employment Rate	-0.065	0.058	2.516***	2.164***	-0.447	0.072	1.261	2.502***
(Province-Level 1997)	(0.352)	(0.325)	(0.751)	(0.789)	(0.394)	(0.329)	(0.843)	(0.831)
Population Density	0.136***	-0.152***	0.499***	0.200**	-0.025	-0.136***	0.189	0.255***
(Province-Level 1997)	(0.038)	(0.049)	(0.056)	(0.088)	(0.066)	(0.046)	(0.116)	(0.089)
Country Effects	Yes	Yes	Nested	Nested	Yes	Yes	Nested	Nested
Country-and-Region E.	No	No	Yes	Yes	No	No	Yes	Yes
Province Effects	No	No	No	No	No	No	No	No
Observations	2,592	2,592	2,199	2,199	2,592	2,592	2,199	2,199
Centered R^2	0.729	0.749	0.584	0.597	0.715	0.741	0.555	0.569
Hansen J Test					0.012	1.678	0.764	1.225
- p-value					0.912	0.432	0.382	0.542
Kleib.-Paap LM Test					28.81	25.07	31.34	25.40
- p-value					0.000	0.000	0.000	0.000
Kleib.-Paap W. F Test					79.48	12.27	27.87	7.564
Exogeneity Test					12.38	5.500	8.893	5.208
- p-value					0.000	0.064	0.003	0.074

† All variables are in natural logs. Heteroskedasticity-robust standard errors (clustered by countries or combinations of countries and Spanish regions) are given in parentheses. *,**,*** denote significance at the 10%, 5%, 1% levels, respectively. The regressions include all countries with at least 630 nationals residing in Spain in the year 1996 (55 origin countries). In columns (e)-(h), the ex-ante co-national network and its interaction with the ex-post transnational network are instrumented with historical emigration flows within Spain (and the corresponding interactions). See section 4.1 for a detailed description of all variables.

In all specifications employed, the estimated coefficient of the ex-ante co-national network variable is positive and statistically significant at the 1% level. In addition, in the specifications of the full model, the estimated interaction effect between the ex-ante co-national network and the ex-post transnational network is positive and always statistically significant at least at the 10% level. The same holds true for the estimated coefficients of the two transnational network variables, the FE 2SLS model of column (h) being the only exception.

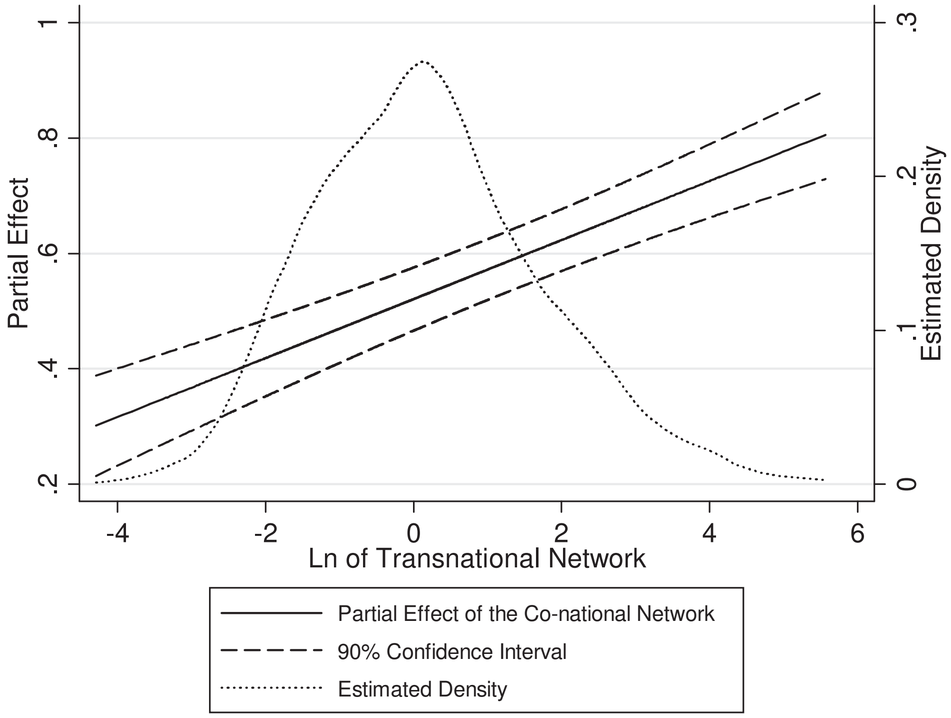
Figure 3 plots the partial effect on follow-up migration due to an increase in the co-national network, $\ln M_{ik}$, against the ex-post transnational network, $\ln(\sum_j \kappa_{ijk})$.[19] It is based on the parameter estimates for β_0 and β_3 reported in column (d) of table 1.[20] The partial effect (straight line) is shown together with the 90% confidence interval (dashed lines) for relevant values of the ex-post transnational network. The figure also includes the estimated density of the ex-post transnational network (dotted line). It reveals that the estimated elasticity is positive and increasing in the size of the ex-post transnational network in the relevant interval. It always lies in the interval between 0.3 and 0.8, the average order of magnitude being roughly consistent with the results obtained in our own earlier work (Neubecker et al., 2012) as well as with the results reported in other studies (Beine et al., 2011). This is quite strong support for the idea that the co-national network exerts an independent positive influence on follow-up migration, but that this influence is the more important, the larger the ex-post transnational network.[21]

[19]The full marginal effect of the co-national network on follow-up migration additionally includes the indirect effects of an increase in the co-national network on follow-up migration via induced changes in the ex-post transnational network.

[20]For expositional reasons, we fix the provincial population size, M_k, at its sample mean in the variable $\sum_j \ln \kappa_{ijk}$.

[21]We have also plotted the effect of the co-national network on follow-up migration as a function of both the ex-post transnational network and the co-national network, accounting for the indirect impact of the co-national network on follow-up migration via induced changes in the ex-post transnational network. For all relevant parameter constellations, the above-described findings go through.

Figure 3: Network Elasticity with Respect to the Co-national Network $\ln M_{ik}$ (Partial Effect)[†]



[†]This figure shows the partial effect of the co-national network for relevant values of the ex-post transnational network. It abstracts from any indirect impact of the co-national network on follow-up migration via induced changes in the ex-post transnational network. It is based on the estimation results from column (d) of table 1. The figure also shows the 90% confidence interval for the partial effect as well as the estimated density of the ex-post transnational network. The population in the destination province, M_k, is fixed at its sample mean.

Our estimates also support the hypothesis that established transnational networks encourage follow-up migration in the same way as co-national networks do (positive and statistically significant estimate of β_1). The estimated coefficient of the ex-post transnational network variable, $\hat{\beta}_2$, is also positive and mostly statistically significant. However, the marginal effect of this variable reads as:

$$\frac{\partial \ln m_{ik}}{\partial \ln \sum_j \kappa_{ijk}} = \beta_2 + \beta_3 \ln \lambda_{iik}. \tag{14}$$

Using the parameter estimates from column (d) of table 1 and evaluating this expression at a zero ex-ante co-national network, this marginal effect is approximately equal to zero.[22] This is an important result in that the ex-post transnational network can only impact on follow-up migration through an existing ex-ante network of co-nationals.

Most coefficient estimates of the control variables have the expected signs. This is especially true for the specifications that eliminate country-and-region fixed effects. Bilateral migrant flows are larger, ceteris paribus, for country-province combinations characterized by a high trading volume or a high inflow of FDI at the regional level; for provinces with a high GDP per capita or a high employment rate; for densely populated provinces, once country-and-region effects are controlled for.

The instruments used in the FE 2SLS estimations reported in columns (e)-(h) of table 1 seem to be essentially valid, relevant, and strong according to various test statistics. In order to test for the validity of the instruments, we perform overidentification tests of all instruments in the form of Hanson J tests. From the reported test values for this statistic, we can never reject the null hypothesis of instrument exogeneity at any reasonable level of confidence. Furthermore, the values of the Kleibergen-Paap LM statistic indicate that our excluded instruments are relevant because we can always reject the null hypothesis of underidentification. The Kleibergen-Paap Wald F test provides some information on the strength of the instruments. This test statistic is above the critical value of 10 in the specifications in which only the ex-ante co-national network is treated as endogenous (columns e and g).[23] This suggests that there is no problem of weak instruments. Following the suggestion by Baum et al. (2007), we compare the values of the Kleibergen-Paap Wald F statistic to the critical values for the Cragg-Donald Wald F statistic provided by Stock & Yogo (2005) in the specifications

[22]For example, for $\lambda_{iik} = \bar{\eta}_{ii}/N_w$, it is equal to -0.06. Recall that we add one to the variable M_{ik}, such that a zero co-national network implies that $\lambda_{iik} = \eta_{ii}/N_w$.

[23]Baum et al. (2007) suggest the use of this test statistic in combination with the "rule of thumb" by Staiger & Stock (1997) to test for weak identification in the case of heteroskedasticity. According to the "rule of thumb", the F statistic should be at least 10 in the case of one endogenous regressor.

in which both variables involving the ex-ante co-national network are instrumented (columns f and h).[24] Based on this comparison, the instruments do not seem to be weak in the specification reported in column (f), yet the same instruments seem to lead to a bias of the FE 2SLS estimator relative to the bias of the OLS estimator of at most 10% in the specification reported in column (h). Based on exogeneity tests of the instrumented regressor(s), we can always reject the hypothesis of exogenous regressors at the 1% or 10% level in table 1.

4.3 Robustness Analysis

The model presented in the previous subsection includes control variables at the province level instead of fixed effects for Spanish provinces. Hence, the estimates reported above might be subject to omitted variables bias. In order to address this issue, we have also estimated a reduced version of our model which does include province fixed effects, but which does not allow us to separately identify the effects of the ex-ante and the ex-post transnational network. The model reads as follows:

$$\begin{aligned} \ln m_{ik} &= \beta_0 \ln\left(\frac{\eta_{ii} M_{ik}}{N_w}\right) + \beta_1 \ln\left(\frac{1}{N_w}\sum_{j\neq i}\eta_{ij}M_{jk}\right) \\ &\quad + \beta_2 \ln\left(\frac{\eta_{ii} M_{ik}}{N_w}\right) \times \ln\left(\frac{1}{N_w}\sum_{j\neq i}\eta_{ij}M_{jk}\right) + \ln\left(N_i - \sum_k M_{ik}\right) + \mu_{ik}. \qquad (15) \end{aligned}$$

Unlike before, the transnational network variable in equation (15) does not include the native population at a given migration destination.

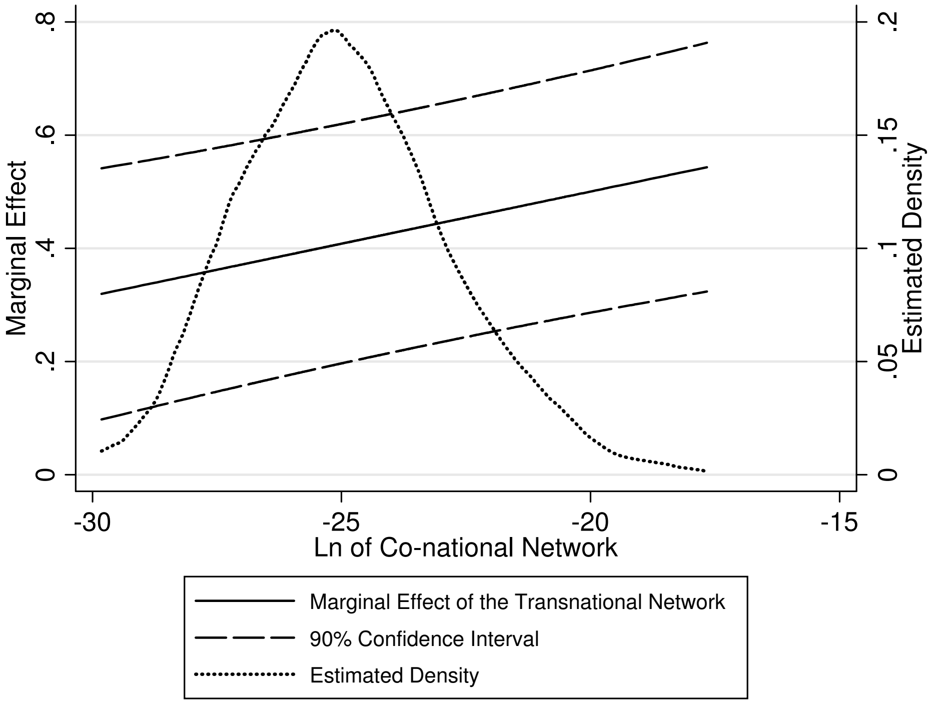
The results from FE and FE 2SLS estimations of the reduced model are reported in table A.3 in the appendix. By and large, they confirm our findings from the estimations of the full model. However, in the FE 2SLS estimations the effects of the two variables involving the transnational network variable are rendered insignificant. This is probably related to the fact that our instrumental

[24] In fact, the Cragg-Donald Wald F statistic is the relevant F statistic in the case that the errors are independent and identically distributed.

variables are weak in columns (f) and (h) of table A.3. A comparison of the Kleibergen-Paap Wald F statistics with the critical values provided by Stock & Yogo (2005) suggests a relative bias of the FE 2SLS estimator of up to 10% and 20%, respectively. Therefore, more attention should be paid to the FE results in columns (b) and (d).

Figure 4 plots the marginal effect of the transnational network for relevant values of the ex-ante co-national network based on the parameter estimates reported in column (d) of table A.3. Again, it also includes the 90% confidence interval as well as the estimated density of the co-national network. The marginal effect of the transnational network on follow-up migration is always positive and increasing in the co-national network, ranging between 0.3 and 0.55 for all observations.

Figure 4: Network Elasticity with Respect to the Transnational Network[†]



[†] This figure shows the marginal effect of the transnational network for relevant values of the co-national network. It is based on the estimation results from the reduced model reported in column (d) of table A.3. The figure also shows the 90% confidence interval for the marginal effect as well as the estimated density of the co-national network.

Furthermore, we have tested the robustness of our main findings using a different empirical measure to proxy cultural and geographical proximity in the network variables considered in model (13). Using a dummy variable indicating whether two countries share a common language instead of the inverse of geographical distance for the construction of the network variables, we obtain results that are qualitatively similar to those reported in table 1.

5 Conclusion

We have argued that the existing literature estimating network effects in migration has been overly restrictive in its definition of migrant networks. It has ignored the role of transnational networks in determining the scale of migration. In this paper, we provide evidence that transnational networks are relevant predictors of international migration flows, both independently and in conjunction with the network of co-national friends. We do so using aggregate migration data from the recent immigration boom to Spain, detailed by origin countries and destination provinces.

The two novel findings of our paper relative to the existing literature are as follows. First, the number of migrants from a given origin country to a certain migration destination is increasing in the number of settled migrants with nationalities of culturally and geographically proximate third countries. Importantly, this holds true even if the migrant network of co-nationals is small or zero. We attribute this finding to the presence of transnational ties, developed among individuals from culturally and geographically proximate countries before pioneer migration takes place. The second novel finding is a pronounced non-linearity in the standard network elasticity estimated in the literature. More precisely, we find that the positive impact on bilateral migration of an origin-country specific migrant network is the larger, the larger the migrant community of culturally similar individuals holding third-country nationalities. An explanation for this result that we offer in this paper is that friendships formed at destination enhance the value of the migrant community for those left behind, and that friendship formation is more likely among individuals with a similar cultural background.

Our paper employs macro-level data on migrant stocks and flows. An obvious drawback of this approach is that actual network ties among individuals are not observed. Although the correlations between aggregate migrant stocks and flows documented in our paper seem to be well in line with the idea of transnational networks, complementary evidence from the micro level would significantly strengthen our interpretation.

References

Åslund, O. (2005). Now and Forever? Initial and Subsequent Location Choices of Immigrants. *Regional Science and Urban Economics*, *35(2)*, 141–165.

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.

Bauer, T., Epstein, G. S., & Gang, I. N. (2009). Measuring Ethnic Linkages among Migrants. *International Journal of Manpower*, *30(1/2)*, 56–69.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007). Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing. *The Stata Journal*, *7(4)*, 465–506.

Beine, M., Docquier, F., & Özden, C. (2011). Diasporas. *Journal of Development Economics*, *95(1)*, 30–41.

Beine, M., Docquier, F., & Özden, C. (2012). Dissecting Network Externalities in International Migration. University of Luxembourg, mimeo.

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

Chau, N. (1997). The Pattern of Migration with Variable Migration Cost. *Journal of Regional Science*, *37(1)*, 35–54.

Clark, X., Hatton, T. J., & Williamson, J. G. (2007). Explaining U.S. Immigration, 1971-1998. *Review of Economics and Statistics*, *89*(2), 359–373.

Davis, B., Stecklov, G., & Winters, P. (2002). Domestic and International Migration from Rural Mexico: Disaggregating the Effects of Network Structure and Composition. *Population Studies*, *56*(3), 291–309.

Dolfin, S., & Genicot, G. (2010). What Do Networks Do? The Role of Networks on Migration and "Coyote" Use. *Review of Development Economics*, *14*(2), 343–359.

Duncan, O. D., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, *20*(2), 210–217.

Hatton, T. J. (1995). A Model of U.K. Emigration, 1870-1913. *Review of Economics and Statistics*, *77*(3), 407–415.

Lewer, J. J., & Van den Berg, H. (2008). A Gravity Model of Immigration. *Economics Letters*, *99(1)*, 164–167.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of International Migration: A Review and Appraisal. *Population and Development Review*, *19(3)*, 431–466.

Neubecker, N., Smolka, M., & Steinbacher, A. (2012). Networks and Selection in International Migration to Spain. University of Tübingen Working Papers in Economics and Finance No. 35.

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.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.

Stark, O. (1999). *Altruism and Beyond, An Economic Analysis of Transfers and Exchanges Within Families and Groups*. Cambridge University Press, Cambridge.

Stark, O., & Wang, Y. Q. (2002). Migration dynamics. *Economics Letters*, *76*(2), 159–164.

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In Andrews, D. W. K and Stock, J. H. (Ed.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, (pp. 80–108). Cambridge University Press.

United Nations (2009). *Trends in International Migrant Stock: The 2008 Revision (United Nations database, POP/DB/MIG/Stock/Rev.2008).* United Nations, Department of Economic and Social Affairs, Population Division.

Appendix

Table A.1: Data Sources[†]

Data	Source
FDI Flows	`http://datainvex.comercio.es/principal_invex.aspx`, accessed on 10/20/2010
Geographical Distance	`http://www.cepii.fr/anglaisgraph/bdd/distances.htm`, accessed on 10/13/2010
Gross Domestic Product	`http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft35%2Fp010&file=inebase&L=1`, accessed on 09/14/2012
Historical Flows of Foreign Nationals	`http://www.ine.es/en/prodyser/micro_varires_en.htm`, accessed on 10/05/2010
Migrant Communities	`http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fe245&file=inebase&L=0`, accessed on 10/07/2010
Migrant Flows	`http://www.ine.es/en/prodyser/micro_varires_en.htm`, accessed on 10/05/2010
National Immigrant Survey 2007	`http://www.ine.es/prodyser/micro_inmigra.htm`, accessed on 10/05/2010
Population	`http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft35%2Fp010&file=inebase&L=1`, accessed on 09/14/2012
Surface Area	`http://www.ine.es/jaxi/tabla.do?path=/t43/a011/a1998/densidad/a2008/l0/&file=t10031.px&type=pcaxis&L=0`, accessed on 09/11/2012.
Trade Flows	`http://datacomex.comercio.es/principal_comex_es.aspx`, accessed on 10/20/2010
(Un-)Employment	`http://www.ine.es/jaxi/menu.do?L=1&type=pcaxis&path=%2Ft22/e308_mnu&file=inebase`, accessed on 10/10/2010
World Population	`http://esa.un.org/unpd/wpp/unpp/panel_population.htm`, accessed on 09/26/2012

[†] See section 4 for the definitions of variables and aggregation rules.

Table A.2: List of the 55 Countries Considered in the Empirical Analysis, by World Regions

EAST ASIA & PACIFIC	Cuba	NORTH AMERICA	WESTERN EUROPE
China	Colombia	& AUSTRALIA	Austria
Japan	Dominican Republic	Australia	Belgium
Korea	Ecuador	Canada	Denmark
Philippines	El Salvador	United States	Finland
EASTERN EUROPE	Honduras	SOUTH ASIA	France
& CENTRAL ASIA	Mexico	India	Germany
Bosnia and Herzegowina	Peru	Pakistan	Ireland
Bulgaria	Uruguay	SUB-SAHARAN	Italy
Poland	Venezuela	AFRICA	Netherlands
Romania	MIDDLE EAST	Angola	Norway
Russia	& NORTH AFRICA	Cape Verde	Portugal
LATIN AMERICA	Algeria	Equatorial Guinea	Sweden
& CARIBBEAN	Egypt	Gamba	Switzerland
Argentina	Iran	Guinea	United Kingdom
Bolivia	Lebanon	Mauritania	
Brazil	Morocco	Senegal	
Chile	Syria		

Table A.3: Estimation Results for the Reduced Model Specification†

	Dependent Variable: Migration Inflow (Province-Level 1997-2006)							
	Fixed Effects				Fixed Effects Two Stage Least Squares			
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Co-national N.	0.682***	1.259***	0.539***	0.931***	0.953***	1.123***	0.825***	0.722***
(Province-Level 1996)	(0.028)	(0.188)	(0.029)	(0.143)	(0.070)	(0.213)	(0.080)	(0.206)
Transn. N.		1.102***		0.869***		0.516***		0.061
(Province-Level 1997)		(0.244)		(0.200)		(0.178)		(0.261)
Co-n. x *Transn. N.*		0.027***		0.018***		0.010		-0.006
(Province-Level 1997)		(0.009)		(0.007)		(0.009)		(0.010)
Trade Flow	0.005	0.011	0.004	0.005	0.004	0.007	0.005	0.006
(Province-Level 1996)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
FDI Flow	0.012**	0.012**			0.004	0.005		
(Region-Level 1997)	(0.005)	(0.005)			(0.005)	(0.005)		
Country Effects	Yes	Yes	Nested	Nested	Yes	Yes	Nested	Nested
Country-and-Region E.	No	No	Yes	Yes	No	No	Yes	Yes
Province Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,592	2,592	2,199	2,199	2,592	2,592	2,199	2,199
Centered R^2	0.792	0.798	0.670	0.674	0.770	0.369	0.635	0.624
Hansen J Test					0.023	1.089	0.379	1.686
- p-value					0.880	0.580	0.538	0.430
Kleib.-Paap LM Test					20.13	11.66	24.27	17.82
- p-value					0.000	0.009	0.000	0.000
Kleib.-Paap W. F Test					30.70	8.622	18.48	7.142
Exogeneity Test					14.29	11.36	11.03	12.17
- p-value					0.000	0.003	0.001	0.002

† All variables are in natural logs. Heteroskedasticity-robust standard errors (clustered by countries or combinations of countries and Spanish regions) are given in parentheses. *,**,*** denote significance at the 10%, 5%, 1% levels, respectively. The regressions include all countries with at least 630 nationals residing in Spain in the year 1996 (55 origin countries). In columns (e)-(h), the co-national network and its interaction with the transnational network are instrumented with historical emigration flows within Spain (and the corresponding interactions). In column (f), all province effects are partialled out in the calculations of the test statistics. See section 4.1 for a detailed description of all variables.