

Publicado en: Small Bus Econ (2018) 51:735-755

<https://doi.org/10.1007/s11187-017-9962-7>

Networks and Self-Employed Migrants

Joan Martín-Montaner, *Universitat Jaume I and IEI*

Guadalupe Serrano-Domingo, *Universitat de València*

Francisco Requena-Silvente, *Universitat de València*

Corresponding autor:

JOAN MARTÍN-MONTANER

Universitat Jaume I and Instituto de Economía Internacional

Av. Vicente Sos Baynat, s/n

12071 – Castelló (Spain)

Tel.: +34 666 20 3292 / +34 964 38 71 65

e-mail: jmartin@uji.es

This work was supported by the Ministerio de Economía y Competitividad [grant numbers ECO2014-58975-P, ECO2015-68057-R] and the Universitat Jaume I [grant number P1.1B2013-22]

ABSTRACT

This paper examines the immigrant's decision to become self-employed. Personal characteristics and attitudes are crucial to this decision, but recent literature also highlights the importance of social networks as providers of information and financial support. To date empirical research dealing with the network effect has considered self-employed migrants as a homogeneous group. In contrast, our paper examines the impact of the size and the composition of geographic concentration of co-nationals on the propensity to be self-employed among the foreign born workers taking into account two possible choices: own-account workers and entrepreneurs. Using micro-level data of migrant workers in Spain over the period 2000-2009, the results show that networks affect positively the probability of becoming self-employed. However, they do not affect the likelihood of becoming an entrepreneur, which is determined by individual characteristics such as education and number of years of residence. Thus, our results suggest policies than enable immigrant assimilation can play a signification role in generating entrepreneurial activity.

Keywords: immigration, self-employment, networks, entrepreneurship, sequential logit

JEL: J21, J22, J61

Networks and Self-Employed Migrants

1. INTRODUCTION

In this paper we investigate the impact of ethnic networks on immigrant workers' decision about becoming self-employed. Although the determinants of the self-employment decision of migrants have been recently analysed, the role of networks has received less attention. Indeed, to the best of our knowledge, the impact of ethnic networks on the choice between becoming own-account worker or entrepreneur remains unexplored.

One stylised fact first pointed out by Borjas (1986) for the U.S. and later supported by Yuengert (1995), Le (2000) and Hammarstedt (2004), for the U.S., Australia and Sweden, respectively, is that the self-employment option increases among immigrants as they live longer in the host country. However, there are important differences among individual countries in terms of rates of self-employment between foreign workers and native-born workers. For example, in 2007-08, the share of self-employment of natives in Greece was 26,5%, whereas in the case of migrants was 10.2%. In contrast, for same period, the shares in Poland were the opposite: 11,2% for natives and 29,4% for migrants (see OECD International Migration Outlook, 2011). Such country differences may have been driven by two factors. First, there could be differences in attitudes to self-employment in the countries of origin of migrants. Pietrobelli et al (2004) and Gindling and Newhouse (2014) confirm the large heterogeneity of self-employment rates across developing countries - the main type of country of origin of migration – but Fairlie and Meyer (1996) shows that the patterns observed at home are not necessarily reproduced after migration to the host country. Second, there are differences in the way migrants accede and are assimilated in the host countries' labour market, leading to differences in self-employment between natives and migrants. One reason for these differences is that, because foreign-born workers are disadvantaged when competing for the same jobs and wages as native-born workers, they necessarily turn to self-employment in order to get an income. Another reason is that the preference of migrants for products originated in their country of origin (the so-called ethnic products) stimulates some co-nationals to produce and commercialize them.

The importance of both explanations is affected by the degree of concentration of migrants in the receiving country in two ways. The higher the number of co-nationals in the host country and the closer they live from each other, the more likely it is that significant demand for ethnic products will appear.

Second, specific attitudes towards self-employment may differ across countries, but this attitude is also affected by the job opportunities they find in the host country. The geographical agglomeration of co-nationals is strongly correlated with the degree of assimilation they have experienced, which in turn affects the job opportunities they enjoy (Chiswick and Miller, 2005). Our focus is on this latter issue as the degree of spatial concentration of migrants introduces additional dimensions in the analysis of the causes of migrant's self-employment. High concentration contributes to the appearance of networks that may allow individuals access to the resources to start a business. However, intra-group competition might attenuate or prevent those positive effects to appear. Thus, for instance, the existence of networks could enhance the recruitment of co-national individuals as employees, so that they dismiss the self-employment option. Therefore, the simple concentration of co-nationals and the existence of a helpful network is not the same: the skill composition of the members of the network is crucial.

In this paper, we identify the existence of geographic migrant networks and then we measure the effect of these networks on the probability to become self-employed in Spain. For that purpose, we exploit individual-level micro data obtained from the Labour Force Survey (LFS) over the period 1999-2009. The data shows an important rise of the migrant population in Spain, as the percentage of foreign-born workers in the LFS rises from 4 percent in 1999 to a maximum of 16 percent in 2009. In contrast to past papers that analysed the migrants' decision in a dichotomous setup (employee/self-employed), we propose a three choice model (employee, own-account worker and employer) and search for the best model to empirically analyse the mechanism through which the decision is made.

Data reveal that the migrants' employment status decision is a sequential process: migrants decide first whether to become employee or self-employed, and, then, whether to become an own-account worker or an entrepreneur. Next, when we examine the determinants of the employment choices, the results show that the migrants' self-employment choice is positively affected by the existence of networks in which their members have been previous self-employment experience in the host country. Personal characteristics such as qualification, the years of residence in the host country, being married and the age also contribute favourably to this decision, whereas women are less likely to become self-employed. However, once we consider the choices between own-account worker and entrepreneur, the positive impact of ethnic networks is only relevant for the own-account choice. In other words, we find that entrepreneurship among migrants

is not enhanced by the existence of networks; only education attainment and the years of residence in the host country affect positively the job creator choice.

The rest of the paper is structured as follows. Section 2 defines the conceptual framework for the analysis and establishes the hypothesis to be tested. Section 3 presents the data and provides some descriptive analysis. Our key variables, econometric model and estimation strategy are displayed in Section 4. We present our main results in Section 5 and Section 6 summarizes the key findings and provides the main conclusions of our research.

2. LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

The determinants of the self-employment decision have been on the spotlight of economics and management research for a long time. In this context, it is commonly acknowledged that this decision presents some specific aspects in the case of migrants, which have been deeply analysed from a sociological and economic view (Light 1972; Borjas 1986). First explanations consider the self-employment migrants' decision of migrants as a way to escape from poor conditions in the labour market in the host country, which could explain a higher propensity towards self-employment observed in some immigrant groups. Light (1984) relies upon the disadvantage theory, highlighting the more limited options enjoyed by migrants due to scarcity of endowments in the host country (i.e. poor communication skills, non-transferable labour skills, discrimination in the labour market). Another explanation is the existence of a sort of self-employment culture transferred from the home country to the host one. However, the empirical evidence is still inconclusive. While Yuengert (1995) and Foreman-Peck and Peng (2013) find a positive link but the strength of entrepreneurial culture is different across ethnic groups or nationalities, Fairlie and Meyer (1996) observe fail to support the "culture" argument. Instead, the share of self-employment in the countries of origin of migrants could be more related to that country's stage of development rather than to a cultural attitude towards self-employment (Pietrobelli et al. 2004; Gindling and Newhouse 2014).

The environment, in which the self-employment decisions are adopted, could be influenced by the existence of ethnic enclaves, as they might affect the migrant's situation in the host country. These ethnic enclaves may enhance immigrant self-employment in various ways. First, they provide valuable information to immigrants about self-employment such as activities of interest, business methods, adding

credibility as a consequence of the relationships of mutual trust between the network members. Second, they facilitate access to resources such as human capital (for example, on-the-job formation in co-ethnic's business) and financial support which can be crucial in contexts of liquidity constraints or limited access to other forms of external funding. Third, migrants could have comparative advantage in the production of goods that are specific to people from a given country or language group. These advantages could appear because of the existence of economies of scale and cheaper costs hiring people from the same groups of reference. Finally, enclaves could have also a positive effect on self-employment when there is a lack of host language proficiency in the ethnic community where the migrant lives, making more difficult her assimilation in the host labour market as employees (Chiswick and Miller 2005).

Nevertheless, the existence of enclaves is not enough to help the migrants in their self-employment decision. Certain level of know-how and expertise within the enclave may be an additional condition to have a positive impact on this decision. It can be necessary to take into account both the size and the skills of each group of migrants to characterize them as an effective network. This idea is confirmed by Edin et al. (2004), which show that ethnic networks, only when mass of high-skilled members achieve certain level, have a positive effect on the likelihood of finding a job and on the wages.

When we distinguish own-account workers from employers, the number of papers that examine the different determinants of each self-employment category is still limited. Cowling et al. (2004), using micro-level data for some areas in the UK, observe that formal education reduces the probability of self-employment but increases the probability of becoming a job creator. Kautonen et al. (2014), for the case of native self-employed in the EU-25, find differences in the characteristics that affect the probability to become job creator or own-account worker. Lassmann and Busch (2015) obtain different results for incorporated and total migrant self-employment in the US on the cost of start-up procedures or the effect of self-employment shares. Clark et al. (2017), for the case of UK, studies the determinants of migrant self-employment, and then for a subgroup of employers, and some differences appear in relation to age, qualifications and years of residence, depending of the ethnic group where migrants belong. This evidence justifies the interest of analysing the determinants of own-account work and employers separately.

Two recent papers, Andersson and Hammarstedt (2012) and Toussaint-Comeau (2012), find that the positive impact of ethnic networks on migrants' self-employment depends on the existence of a high rate of self-employment in the ethnic community at which the migrant belongs. However, they do not

consider that self-employment status encompasses different situations. Typically, own-account workers are individuals that have arrived to self-employment either because the difficulties they have to get a wage-paid job or because they are individuals with a preference for the flexibility and freedom acquired by being their own boss (Earle and Savoka 2000). In contrast, job creators are individuals more committed to the entrepreneurial activity (Kautonen et al. 2014) and more risk seeking (Lassman and Busch 2015). Job satisfaction is generally higher for self-employed than for employees, but this satisfaction is even higher for employers than for own-account workers (Blanchflower 2000; Gindling and Newhouse 2014). Given the higher requirements in terms of information and resources in the case of the employer option (Cowling et al. 2004), we expect a higher impact of networks for this option. To the best of our knowledge, no empirical research has dealt with these topics up to this moment.

Thus, in the light of the empirical literature mentioned above on the determinants of self-employment, we formulate two main hypotheses:

H (1): Ethnic networks have a positive effect on the probability of immigrants' self-employment choice.

Considering two types of self-employment (own-account worker and entrepreneurship) and the fact that empirical research highlights the different impact of personal characteristics on the probability of different self-employment status, our second hypothesis is:

H (2): Positive networks effects are not equally determining the probability of immigrants' choice between own-account workers and entrepreneurs.

Although our main interest consists in the two hypotheses above, we must also control for the personal characteristics of migrants. Some personal characteristics may stimulate migrants to enter the labour market through self-employment. Empirical research has shown that the age of the individual, the marital status and the years of residence in the host country tend to increase the probability of self-employment. Thus, older individuals have (on average) more resources to start the business activity and a stronger desire for flexibility. However, this positive impact can be qualified depending on the type of self-employment: For potential entrepreneurs an inverse U-shaped link could be expected (Carroll et al. 2000; Kautonen et al. 2014). Married migrants usually have more resources than single individuals, and may benefit from lower costs of the self-employment activity because help from the spouse (Borjas 1986). This

impact is probably more relevant for own-account workers. The empirical literature has systematically found a lower propensity to self-employment in the case of women (Simoes et al., 2016).

Previous research reveals that the impact of migrants' education level on the type of self-employment is unclear. On the one hand, higher education increases the chance of obtaining a well-paid job as employee, reducing therefore the incentives of self-employment. On the other hand, formal education generates some skills such as analytical abilities or communication skills that are needed to run a business (Simoes et al. 2016). Finally, as immigrants stay more years in the host country, they acquire more knowledge about how institutions work. They may also have the opportunity of accumulating savings to support the setting up of the business. This accumulation of human, physical and financial capital requires time, thus the number of years of residence is likely to increase the probability of both types of self-employment (Clark et al. 2017).

According to the ideas presented above, our analysis of migrant's self-employment decision builds upon the random utility maximization (RUM) approach. Since McFadden (1974), the RUM framework provides the bases for the empirical analysis of discrete choice models, where utility maximization individuals select a particular alternative from a distinct set. Thus, this approach allows us testing not only our previous hypotheses on the determinants of self-employment decision, considering own-account work and entrepreneurship as separate alternatives, but also the mechanisms through which individuals' decisions are taken.

3. DATA DESCRIPTION

Our data come from the Spanish Labour Force Survey (LFS), a sample survey aimed to collect individual information in relation to labour market issues. Individuals fill the survey quarterly, providing not only information about personal issues such as their marital status, age and (very important) country of birth and nationality but also their labour status (being unemployed, self-employed or employee), their occupation, sector of activity and educational attainment. In the case of immigrants, it also includes the number of years since their arrival to Spain.

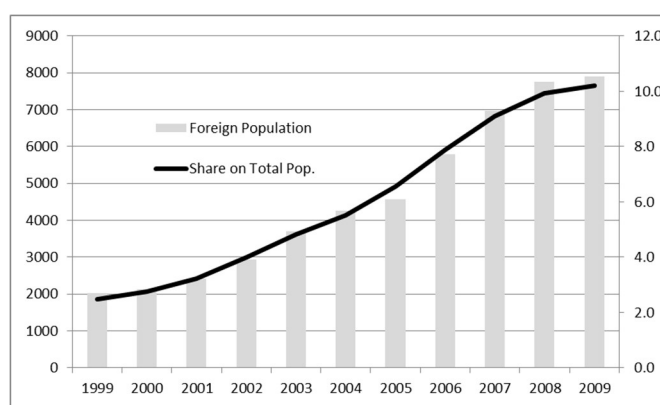
The basic unit is the household. This means that those members of the family who are between 16 and 64 years old and happen to be at home are the ones that fulfil the survey. In other words, respondents will not necessarily be the same each quarter, although the family house stays in the sample. Respondents

change in each wave so do the countries of origin of the migrants. There is a list of 130 nationalities included in the survey over the period 1999-2009 (see Appendix A.3). The number of countries of origin in the sample ranges between 81 in 1999 and 104 in 2009. Thus, our data lacks of a panel structure, as we are not able to follow individuals across time.

Data is referred to each year's second quarter and covers the period 1999-2009, a period of fast, massive and continuous arrival of foreign population to Spain. This trend is truncated after 2010 and unemployment rose rapidly, being the immigrants one of the most affected social groups.¹

As we mentioned before, the foreign-born residents in Spain grew extensively over the 2000-decade. Using LFS individual-level data, Figure 1 shows a huge increase that occurs both in absolute numbers and in relative terms, since working-age foreign-born population resident in Spain multiplied by 5 over the decade, from a 2.5% in 1999 to 10.2% in 2009. Figure 2 shows that the migrants' origin countries in 1999 are mainly EU members, Latin American and Maghreb countries in descending order.

Figure 1. Foreign-Born Individuals in the survey



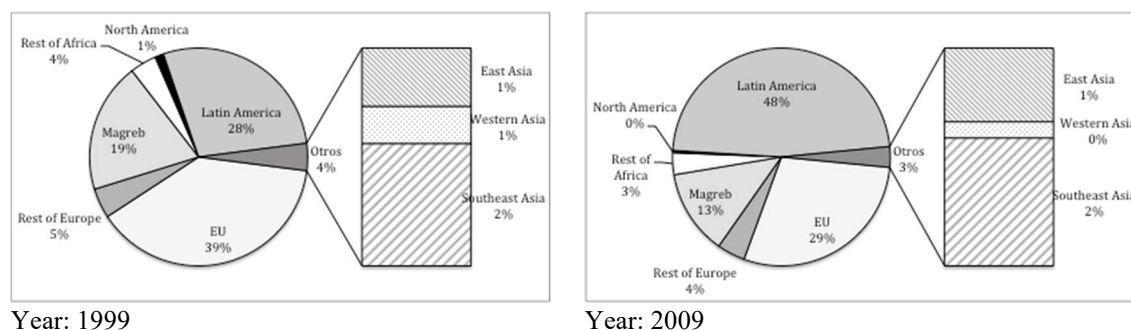
Source: Data from the Spanish Labour Force Survey

These regions remain the main migration sources, but in the last years they exchange their relative size in terms of foreign-born residents in Spain. Latin American migrants account for nearly half the total foreign-born population in 2009 while the relative size of foreign-born communities that come from EU or

¹ The year 2009 was the last year with an increase in the number of foreign population and a positive foreign migration balance in Spain (INE 2014).

Maghreb countries decrease over the decade. It is in the last years in our sample, 2008 and 2009, when the number of countries included in the survey increase significantly, from 81 in 1999 up to 104 in 2009.

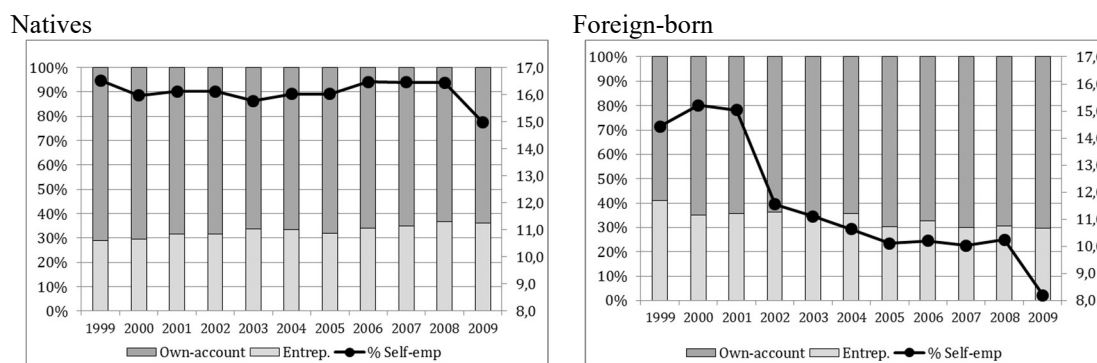
Figure 2. Composition of the Survey by Regional Shares on Foreign-Born Population (%)



Source: Data from the Spanish Labour Force Survey

These changes in the origin of the foreign-born population coexist with a drop of the self-employment share of migrants. It is quite close to that of native's in the first years, but it falls very quickly to 11%-10% in 2002, showing a smoother decline from that year on, see Figure 3. The natives' share, however, remains stable over the years around 16% until 2009, experiencing a 1% decrease in that year.²

Figure 3. Self-employment rates



Source: Data from the Spanish Labour Force Survey

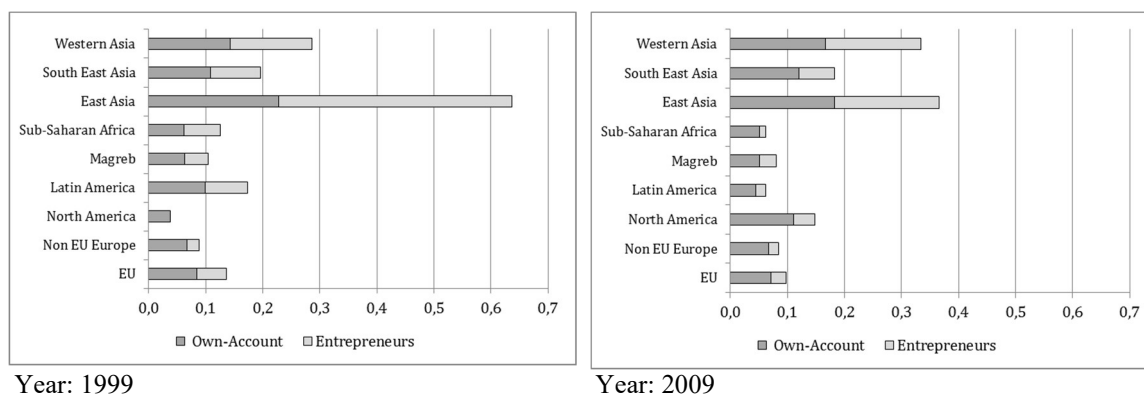
This asymmetric evolution implies an increasing gap between native and foreign self-employment shares from 4.6% in 2002 to 6.8% in 2009. In this case, the beginning of the arrival of the wave of immigrants to Spain in 2002 and the economic boom in this decade may suggest that immigrants in general

² Constant and Zimmerman (2006) also find higher self-employment rates for native workers than for foreign-born workers in Germany. This is not a general result across the EU. According to OECD/The European Commission (2014), the self-employment rates in 2012 for foreign-born people were lower than for native-born people in 11 EU member states.

preferred employee positions and did not choose self-employment in those years as an alternative to unemployment.³

Both for natives and migrants, own-account workers represent the majority of self-employment, although once again we observe a divergent trend for both groups. Thus, the share of entrepreneurs slightly increases across years, whereas it decreases for migrants, more noticeably at the end of the decade.

Figure 4. Self-employment shares by origin area



Year: 1999

Year: 2009

Source: Data from the Spanish Labour Force survey

This result hides a high degree of heterogeneity when the origin of migrants is taken into account. Several differences are observed in Figure 4 both across origin areas and over the time. On the one hand, Asian foreign-born residents are those who show a higher propensity to self-employment and entrepreneurship in 1999, and it remains the highest in 2009. In fact, they show higher self-employment rates than Spanish workers. Similarly, North American migrants that scarcely show any propensity to self-employment in 1999 are the second group more prone to self-employment in 2009. On the other hand, immigrants like Latin American, who are the largest group of foreign-born workers in Spain in 2009, do not show a high propensity for self-employment in this year, more if we compare it to that they show in 1999, and even less for entrepreneurship.

The time span we are considering allow the analysis of the self-employment choice of foreign-born in a context where the global financial crisis has not hit actually hard the Spanish labour market since in the initial years, 2008 and 2009, expansive demand policies were instrumented. Thus, we do not

³ The significant drop in the migrants' self-employment rate in 2002 is due to a change in the composition of immigration by country of origin in the survey and, in particular, for the substantial increase in the number of respondents coming from Latin America and the fall in the respondents coming from East Asia.

consider self-employment as an alternative for unemployment for foreign-born population, but an election they made according to personal and host economy characteristics, where the previous experience of foreign-born from their origin countries could play a key role.

Under this point of view, our objective is to determine up to what extent this migrants' election responds to their ability to create networks. For analytical purposes, we focus on the number of immigrants who are working as self-employed or employees each year -this excludes unemployed immigrants. Furthermore, we will distinguish among own-account or entrepreneur self-employed immigrants since their implications in terms of job creation and job-market policies are not the same.

4. AN ECONOMETRIC MODEL FOR MIGRANTS' SELF-EMPLOYMENT

4.1. *Selection of the econometric model*

The design of our empirical model requires to define the alternative labour status categories for the immigrant workers and the mechanisms through which individual decisions are taken. Individuals face the decision of being hired by a firm or becoming self-employed. Our econometric strategy takes into account two categories of self-employment: own-account workers and entrepreneurs/job creators. Our first aim is to find the adequate econometric model specification. For that purpose, we need to evaluate whether each category of self-employment constitutes a qualitatively separated and independent category.

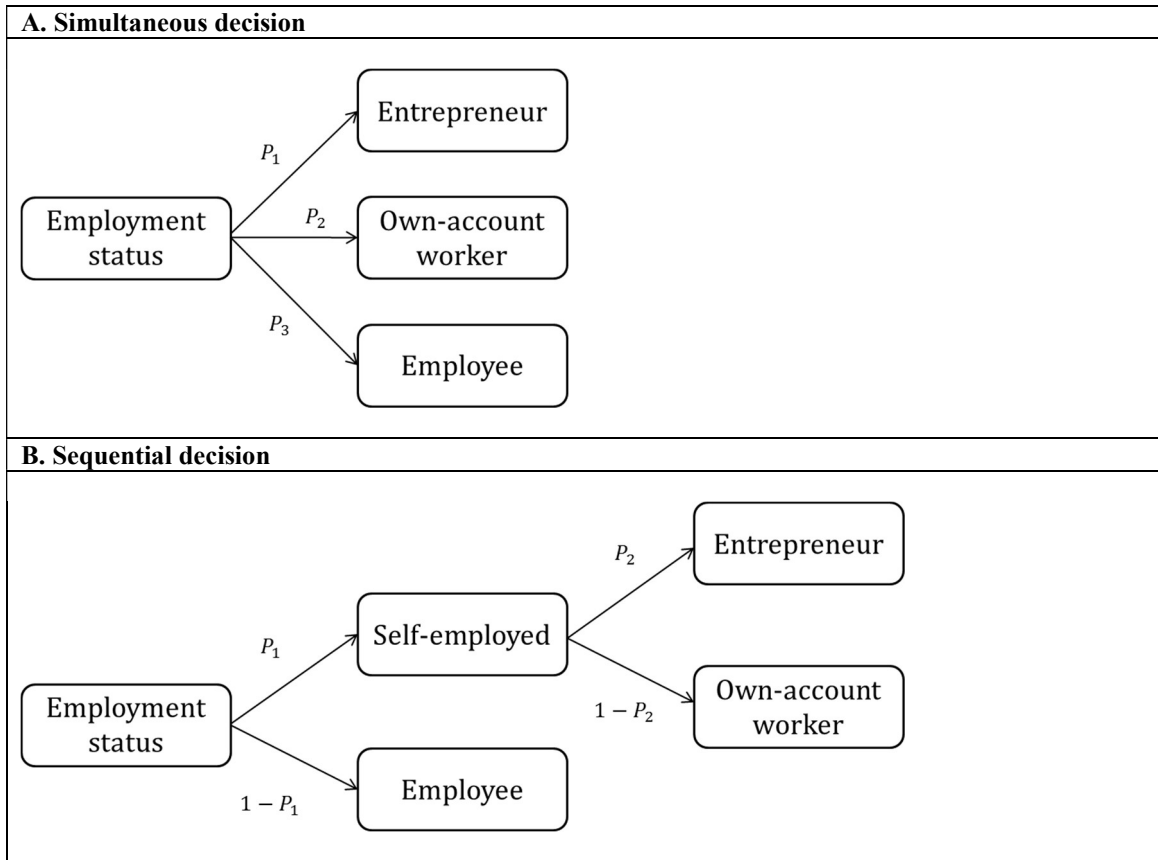
Let us consider an immigrant i , belonging to the ethnic (nationality or language) group k , living in the area j , who faces a choice m among three different employment status: *employee*, *own-account worker* or *employer*. We are interested in the determinants affecting the probability that individual ijk chooses a particular alternative m .

Our first specification considers that individuals face simultaneously three independent options (see Figure 5.A). For that purpose, we use a multinomial logit model (Greene, 2012) such that the probability that individual ijk choose alternative m is:

$$P_{m,ijk} = \Pr(Y_{ijk} = m) = \frac{\exp(\beta_0 + \sum_h \beta_{m,h} X_{hi} + \alpha_{m,j} + \gamma_{m,k})}{\sum_{l=1}^3 \exp(\beta_0 + \sum_h \beta_{l,h} X_{hi} + \alpha_{l,j} + \gamma_{l,k})}$$

where the model fits $0 < P_{m,ijk} < 1$ and $\sum_{i=1}^3 P_{l,ijk} = 1$. The vector X_i includes a set of personal characteristics of individual i , and fixed effects are included for immigrant's area of residence, α_j , and region of origin, γ_k . To ensure that the model is identified, a base category is established by setting $\beta_3 = 0$. In our analysis, the base outcome is *employee*, and the other two alternatives are own-account worker and entrepreneur.

Figure 5. Migrant decisions about employment status.



In our second specification self-employed workers decide between being own-account workers and being entrepreneurs after having dismissed the employee option (see Figure 5.B). This implies that there are two stages in the decision: in the first stage, migrants consider employee and self-employed status as substitutive alternatives, regardless of the self-employment status. In the second stage, they choose between the two self-employment alternative statuses: own-account worker and entrepreneur. The reason underlying

this pattern of decisions is that at the first stage, an individual does not know the values of the utilities of the second-stage alternatives as such knowledge is costly or simply not available at the first stage. Thus, the values of the utilities of the second-stage alternatives do not determine the choice at the first stage, and, consequently, factors determining the election in the first transition, or the impact of such factors in each transition, may be different.

We use a sequential logit (Tutz, 1991) to model this transition of binary decisions, which implies that choices are not entirely independent, because the choice of one of the self-employment categories implies a previous choice to be made.⁴ The sequential logit model allows the analysis of how covariates determine these two transitions probabilities. This is done by estimating a logistic regression for each transition in the subsample that is at risk. All the individuals in the sample face transition 1:

$$P_{1,ijk} = Pr(Y_{ijk} \in \{own - account, entrepreneur\}) \quad (1)$$

$$P_{1,ijk} = \Lambda \left(\beta_{01} + \sum_h \beta_{h1} X_{hi} + \alpha_{j1} + \gamma_{k1} \right)$$

And if passing transition 1, individual face a new decision:

$$P_{2,ijk} = Pr(Y_{ijk} \in \{entrepreneur\} \mid Y_{ijk} \in \{own - account, entrepreneur\}) \quad (2)$$

$$P_{2,ijk} = \Lambda \left(\beta_{02} + \sum_h \beta_{h2} X_{hi} + \alpha_{j2} + \gamma_{k2} \right)$$

The likelihood function for an individual i , group k and area j can be written as:

$$L_{ijk} = \begin{cases} 1 - P_{1,ijk} & \text{if } Y_{ijk} = Employee \\ P_{1,ijk}(1 - P_{2,ijk}) & \text{if } Y_{ijk} = Own - account worker \\ (P_{1,ijk})(P_{2,ijk}) & \text{if } Y_{ijk} = Entrepreneur \end{cases}$$

where $P_{1,ijk}$ and $P_{2,ijk}$ are replaced by the corresponding equations (1) and (2). Maximizing this function with respect to the parameters gives the maximum likelihood estimator.

After estimating both econometric models, we need a criterion to determine which one fits better our data. One procedure is to check whether the multinomial estimates verify the assumption of

⁴ Nagakura and Kobayashi (2009) show that the sequential logit model is a limiting case of the nested logit model, in which the values of the utilities of the second-stage alternatives do not affect the decisions at the first stage.

Independence of Irrelevant Alternatives (IIA), which implies that the ratio of choice probabilities for any two alternatives m, n does not depend on characteristics of a third alternative, l . To check whether the IIA assumption holds, we will use the Hausman and Mc Fadden (1984) test -also known as Hausman test .⁵

4.2. Determinants of the migrants' self-employment

In this subsection we present the explanatory variables of our empirical model. First we introduce the measurement of the size and quality of the ethnic network. Second we present the battery of personal attributes that may affect the decision about the two possible self-employment choices.

4.2.1. Measurement of networks

A network will be defined as people having a common attribute (i.e. nationality or country of birth) and sharing information due to geographic proximity. Thus, measuring networks requires taking into account both the size of the group (and the number of links that can be established among people) and the type of information being shared. According to this approach, we use an indicator that combines (i) the relative agglomeration of group k in the area j and (ii) the likeness of those individuals sharing information and/or resources associated to the self-employment decision. Such knowledge and attitude towards self-employment is measured by the average rate of self-employment of group k . Therefore, a higher average of self-employed members in the group could imply a more supportive environment for those about to become self-employed. Thus, we use the measure of network proposed by Bertrand et al (2000):⁶

$$Network_{jk} = \left(\frac{N_{jk}/N_j}{N_k/N} \right) x (\overline{self - employed}_k) \quad (3)$$

where N as the number of migrants in the corresponding group and area of reference.

⁵ The IIA assumption relies on the independence of error terms. When this assumption does not hold – because two or more alternatives are substitutes, or there is correlation among error terms or the effect of unobserved factors- the estimated coefficients will not be consistent.

⁶ In the paper, the authors discuss the advantages and limitations of the indicator. We refer there the interested reader.

In all the specifications we will also add the share of individuals belonging to the same ethnic group living in a specific area relative to the size of that group on the total population ($Enclave_{jk}$) as an additional covariate:

$$Enclave_{jk} = \left(\frac{N_{jk}/N_j}{N_k/N} \right)$$

where k refers to the ethnic group and j to the area. In both cases, in order to obtain the corresponding number of total immigrants in each group and area of destination, we use INE's 2005 methodology elevation factors. These factors capture the representativeness of each individual in our sample in total population, allowing the extrapolation of population aggregates from our individual sample.

This variable controls for omitted personal characteristics correlated with the concentration of the ethnic group in the area –individual's specific characteristics correlated with his decision of living among his own co-nationals community.⁷ Thus, by including the enclave control, it is possible to disentangle the impact of such characteristics on the probability of living in an enclave from the impact of such characteristics on the probability of being self-employed, that is, the network effect.

We need to define both the k group of reference and the j area where they locate. We make use of the maximum level of disaggregation in our database. With regard to the ethnic group k , we use the country of birth of migrants; for the host area j , we use the area of residence defined at NUTS 3 classification (50 provinces in the case of Spain).⁸

4.2.2. Personal characteristics

⁷ Empirical literature suggests that co-nationals tend to concentrate in the same area (Borjas, 1999; Bartel, 1989) resulting in a reverse causality problem from self-employment to ethnic network that may lead to confusing results. For example, one immigrant that decided to become self-employed could choose to locate in one area because of the high concentration of co-nationals in it, but not because the community has the knowledge and means to facilitate him becoming self-employed.

⁸ Cueto et al. (2015) also use the Spanish LFS to analyse the determinants of migrant's self-employment over the period 2005-2011, though they do not investigate the role of social networks. They use nine world regions to identify the origin of migrants instead of countries of birth. The measurement of enclaves is done by means of the share of broadly-defined groups of migrants in the province (NUTS3).

The decision of becoming self-employed, as any personal decision, depends to a great extent on the own specific characteristics of the individual making the decision (Yuegert, 1995; Lee, 2000; Hammarstedt, 2004; Andersson and Hammarstedt, 2015; Toussaint-Comeau, 2012). Following previous empirical literature, and depending on availability in our database, we include a battery of personal characteristics such as gender, marital status, age, years of residence in the host country and educational attainment. Data comes from the LFS micro data over the period analysed. Table 1 provides a basic description of the covariates included in the regression analysis. We split continuous variables such as years of residence and years of education into intervals in order to identify clearly differences in the impact of these indicators on the decision of self-employment.

Table 1. Migrants' personal characteristics.

<i>Personal characteristics</i>	<i>Definition</i>
Age_{ijk}	Age of each individual
$Female_{ijk}$	Takes value 1 if the individual is a woman and zero otherwise
$Married_{ijk}$	Takes value 1 if the individual is married and zero otherwise
Years of residence	
$YR_{ijk} \leq 1$	Takes value 1 if the individual dwells in the host country up to a year (omitted)
$1 < YR_{ijk} \leq 5$	Takes value 1 if the individual dwells in the host country more than a year up to five
$5 < YR_{ijk} \leq 10$	Takes value 1 if the individual dwells in the host country more than five years up to ten
$YR_{ijk} > 10$	Takes value 1 if the individual dwells in the host country more than ten years
Educational attainment	
$Edu0_{ijk}$	Takes value 1 if the individual did not complete compulsory or primary education and zero otherwise (omitted)
$Edu1_{ijk}$	Takes value 1 if the individual completed compulsory or primary education and zero otherwise
$Edu2_{ijk}$	Takes value 1 if the individual completed secondary education and zero otherwise
$Edu3_{ijk}$	Takes value 1 if the individual completed university or upper studies and zero otherwise

Notes: A brief mention is due for the educational attainment measurement: the survey includes a highly disaggregated classification following UNESCO's International Standard Classification of Education (ISCED, 1997 version). All the items included have been aggregated up to four levels: no education, primary, secondary and tertiary education.

5. RESULTS

5.1. Main results

The empirical analysis is based on the estimation of a model where the probability of immigrants choosing among the alternatives for job status depends on the existence of ethnic networks and on individual's characteristics. Together with these covariates, we also include several controls. We use fixed effects to control for specific time-invariant characteristics of the immigrant's location area within the host country, sector fixed effects controlling for specific characteristics of each sector of economic activity in which the immigrant works, and year fixed effects.⁹ Moreover, immigrant's world regions of origin fixed effects are also included to control for specific characteristics across co-nationals or ethnic groups that may affect immigrant's choice.

Our set of choice includes three alternatives. As discussed in section 3, we can study the determinants of this decision through two alternative econometric model specifications. Multinomial logit models are appropriate if immigrants face all the three options simultaneously, whereas a sequential logit model would be our choice if they face a two-phase decision. To solve the dilemma we test the validity of the multinomial model through the Hausman test. In our case, the IIA assumption may hold if the individual considers only two employment alternatives without considering the third one (for example, considering Employee versus Own-account employment and ignoring the possibility of Entrepreneurship). The null hypothesis of the odds ratio being independent of the consideration of a third outcome in employment status is tested by comparing the full multinomial logit model estimation for all three alternatives and the restricted estimation in which one alternative is omitted.

The outcome for the Hausman test is shown in Table 2. The IIA assumption is rejected regardless of the omitted employment position. Therefore, the multinomial logit specification is not adequate in our case since the choice between any two alternatives is not independent of the third one. We turn therefore to the sequential logit specification. The outcome for the multinomial logit specification is reported in the Appendix A.

Table 2. Hausman test for IIA assumption

⁹ One potential problem that may arise from this specification is the impossibility of identifying the ethnic network effects on the probability of choosing the self-employment alternative because area-specific or ethnic groups' unobserved characteristics might cause higher self-employment rates in some provinces or some ethnic communities.

Omitted category	MLE2	Evidence
Employees	66.74 [0.097]	Against H_0
Own-account workers	113.07 [0.000]	Against H_0
Entrepreneurs	69.38 [0.064]	Against H_0
χ^2 Degrees of freedom	53	

H_0 : Independence of alternative labour market positions. *p-values* in brackets.

We estimate therefore a two-phase transition model (sequential logit model) whose results are shown in Table 3. The two stages or transitions in the immigrant's decision are labelled Transition I and Transition II, respectively.¹⁰ All the estimates are displayed in terms of odds ratio.¹¹

Let us focus first on the decision between becoming an employee or self-employed. In Table 3, the network indicator is larger than 1 and statistically significant, so the probability of self-employment increases with the value of our network indicator. A higher value is achieved through a higher expertise in self-employment (inferred from a higher average of self-employment share of co-ethnics in the host country) for a given density of co-ethnic population in the region.

¹⁰ A sequential logit model can be estimated quite simply by estimating a number of logit models. Nevertheless using Stata's `seqlogit` package also allows for hypotheses testing across transitions since the entire model is estimated simultaneously and for a sensitivity analysis to investigate the potential influence of unobserved variables, see Buis (2011).

¹¹ The odds ratio ($\exp[\beta]$) is the effect of the independent variable on the probability of the event divided by the probability of the non-event. Thus, positive (negative) coefficients lead to odds ratios higher (less) than one, implying the event is more (less) likely to occur. Odds ratios equal to 1 mean that there is a 50/50 chance that the event will occur.

Table 3. Sequential Logit Specification: Baseline model

Trans I <i>Reference category: Employee.</i>		
Trans II <i>Reference category: Own-account worker</i>		
VARIABLES	Transition I	Transition II
<i>Network_{jk}</i>	1.005*** (0.0007)	0.999 (0.0002)
<i>Age_{ijk}</i>	1.057*** (0.0134)	1.035 (0.0262)
<i>Age²_{ijk}</i>	0.999 (0.0002)	0.999 (0.0003)
<i>Female_{ijk}</i>	0.474*** (0.0205)	0.906 (0.0752)
<i>Married_{ijk}</i>	1.409*** (0.0647)	1.092 (0.1007)
<i>Years of residence</i>		
1 < <i>YR_{ijk}</i> ≤ 5	1.248*** (0.1061)	1.372 (0.2860)
5 < <i>YR_{ijk}</i> ≤ 10	1.994*** (0.1733)	2.161*** (0.4439)
<i>YR_{ijk}</i> > 10	2.779*** (0.2512)	2.637*** (0.5460)
<i>Educational attainment</i>		
Primary _{ijk}	1.002 (0.0945)	1.593** (0.3326)
Secondary _{ijk}	1.395*** (0.1339)	1.517** (0.3246)
Tertiary _{ijk}	1.542*** (0.1496)	1.752*** (0.3752)
<i>Enclave_{jk}</i>	0.887*** (0.0155)	1.013 (0.0089)
<i>Constant</i>	0.015*** (0.0049)	0.065*** (0.0439)
Akaike IC	23241.95	
Observations	32,638	

Note: All estimates in odd ratios. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Fixed effects: (1) Area of residence: *Spanish NUTS2*. (2) Area of origin: *EU (3 groups), Non-EU Europe, Maghreb, Sub-Saharan Africa, North America, Central America and the Caribbean, South America, East Asia, Western Asia, Southeast Asia and Oceania*]. (3) Sectors: *Agriculture, Manufacturing and Services*. (4) Years: 1999 to 2009.

With regard to the outcome for personal characteristics, we confirm that self-employment is less likely for females relative to males and more likely for married individuals. In addition, the probability of self-employment increases linearly with the age of the migrant: the fact that older workers may have higher experience and easier access to financial resources could explain this outcome. In a similar fashion, more years of residence – associated to higher assimilation and knowledge of the host country - has a positive impact on the self-employed status probability. Finally, we do not obtain significant statistical differences between being primary educated or not being educated at all (our omitted category), whereas both secondary and tertiary education increase the probability of self-employment.

In the second transition, the outcome changes substantially. Networks become no significant. The lack of impact also appears in three of the personal characteristics: age, gender and marital status. With regard to the years of residence, they are relevant on the probability of becoming a job creator after having spent more than five years in the host country. Finally, all categories of educational attainment are significant in the second transition.

Thus, networks are relevant for the self-employment decision but not for the entrepreneurship decision, in which other factors such as immigrant's education or years of residence will outweigh its influence. In other words, a migrant who has chosen self-employment is more likely to become an entrepreneur because of the personal knowledge and experience acquired in the host country rather than because of the information and resources provided by networks. When individuals decide becoming job creators, they can face the financial and informational requirements without relying on networks.

We test in Table 4 the homogeneity of the estimates of the parameters for the years of residence and the educational attainment, both within each transition and between transitions. First, we analyse if the three parameters corresponding to the intervals of the years of residence are equal $[\beta_{1,yr5} = \beta_{1,yr10} = \beta_{1,yr(+10)}]$; as the null hypothesis is rejected, we can confirm that the effect of the number of years in the host country increases as years go by. The same conclusion can be achieved for the Transition II with regard to the two significant parameters $[\beta_{2,yr10} = \beta_{2,yr(+10)}]$. However, the effects of the number of years on the decision of become self-employed and the decision of become a job creator are not statistically different: we also

test for the equality of the parameter of $5 < YR_{ijk} \leq 10$ across transitions $[\beta_{1,yr10} = \beta_{2,yr10}]$, and the same for $YR_{ijk} > 10$ $[\beta_{1,yr(+10)} = \beta_{2,yr(+10)}]$; in both cases the null hypothesis is accepted.

With regard to the role of educational attainment, the Wald test indicates that the impact of each educational category on the dependent variable is not statistically different in most of the estimates. In Transition I the null hypothesis of equality of the parameters $[\beta_{1,sec} = \beta_{1,ter}]$ is rejected at 5%, but in Transition II the impact of the different educational levels $[\beta_{2,prim} = \beta_{2,sec} = \beta_{2,ter}]$ is not statistically different. Across transitions, the equality of the estimates for the secondary and tertiary educational levels is not rejected.

Table 4. Wald tests for equality of estimated parameters.

Years of residence					
<i>Inside transitions</i>			<i>Between transitions</i>		
$\beta_{1,yr5} = \beta_{1,yr10} = \beta_{1,yr(+10)}$	167.47	[0.000]	$\beta_{1,yr10} = \beta_{2,yr10}$	0.13	[0.721]
$\beta_{2,yr10} = \beta_{2,yr(+10)}$	4.40	[0.036]	$\beta_{1,yr(+10)} = \beta_{2,yr(+10)}$	0.05	[0.818]
Educational attainment					
<i>Inside transitions</i>			<i>Between transitions</i>		
$\beta_{1,sec} = \beta_{1,ter}$	4.08	[0.043]	$\beta_{1,sec} = \beta_{2,sec}$	0.13	[0.716]
$\beta_{2,prim} = \beta_{2,sec} = \beta_{2,ter}$	2.46	[0.292]	$\beta_{1,ter} = \beta_{2,ter}$	0.31	[0.580]

Note: Based on estimates displayed in Table 3. *p-values* in brackets.

Thus, our results point out the relevance of educational attainment on self-employment probability, and the relevant and higher effect of the years of residence, once the immigrant reaches a certain degree of assimilation within the host economy.

The previous results based on the sequential logit could be biased due to the potential influence of unobserved heterogeneity even though the unobserved factors are not confounding variables (Cameron and Heckman, 1998; Buis 2011). Unobserved factors may lead to bias problems through two mechanisms. The omission of an unobserved variable in this model implies estimating the probability of passing a transition averaged over such omitted variable rather than the original one and, given model's nonlinearities, the impact of observed covariates on the averaged probability differ from their impact on probability -the averaging mechanism. Additionally, the omitted variable could become a confounding factor in the next transition since individuals deciding on that stage are a sub-sample of the initial sample which may lead to

biased estimates -the selection mechanism. Thus, following Buis (2010, 2011), we perform a sensitivity analysis of our estimates for the impact of network effects on the probability of immigrant's self-employment status choice -without unobserved heterogeneity- to different amounts of unobserved heterogeneity. We specify a number of plausible scenarios concerning the distribution of a hypothetical unobserved and standardised variable and re-estimate our model given such scenarios to assess the relevance of unobserved heterogeneity in our analysis.¹² This is done in Appendix B. The sensitivity analysis shown in Table B1 brings the conclusion that unobserved heterogeneity is not a serious problem in our estimates since the results are robust to any of the assumed unobserved heterogeneity effects in the first transition. In the second transition, results are robust up to the highest impact of unobserved heterogeneity assumed as reasonable in the analysis. In this latter case, extreme values for the impact only affect the significance of the coefficients but do not substantially change their sign or magnitude.

5.2. Robustness Analysis

The main conclusions stemming from the sequential logit model presented in Table 3 are very robust to alternative specifications. In the present section we analyse the sensitivity of the estimates to different definitions of the groups of reference of migrants and their area of residence in the host country. We also examine whether our results are affected by the period analysed and the number of sending countries included.

5.2.1. Alternative definitions of the reference framework for networks

As a robustness check for the sequential logit results, we will consider additional definitions of the national or cultural groups that could result in the existence of networks. To do so, we consider two alternative groups of reference defined by (1) countries sharing the same official language or (2) world regions defined by geographical and/or cultural proximity of countries. The list of groups of countries by language and world regions groups can be found in Appendix A.3. According to our data availability, we

¹² Unobserved heterogeneity is considered as a weighted sum of all unobserved variables that should be controlled for. We assume the unobserved variable normally distributed with constant value and effects over transitions (Buis, 2011).

can also consider a lower disaggregation of areas of residence. We will consider an alternative definition for such area of residence, corresponding to the NUTS2 regional classification (*Comunidades Autónomas*, NUTS2) to check the robustness of our results. The new set of estimates is displayed in Table 5.¹³ They also include our estimates from Table 2 for the sake of comparison. The latter estimation remains our preferred model according to the Akaike criterion.

Table 5. Sequential Logit Estimates. Robustness analysis

Definition of ethnic groups and area of residence

Transition I *Reference category: Employee.*

Transition II *Reference category: Own-account worker.*

A1: NUTS2			A2: NUTS3	
GI: By Country	Trans I	Trans II	Trans I	Trans II
$Network_{jk}$	1.002 (0.0013)	0.998 (0.170)	1.005*** (0.0007)	0.999 (0.0002)
AIC Criterion	23435.49		23241.95	
G2: By Language Group	Trans I	Trans II	Trans I	Trans II
$Network_{jk}$	1.288* (0.1869)	0.583 (0.2325)	1.631*** (0.1331)	0.995 (0.0184)
AIC Criterion	23469.63		23286.68	
G3: By Region	Trans I	Trans II	Trans I	Trans II
$Network_{jk}$	1.517** (0.2528)	1.315 (0.2244)	1.078*** (0.2527)	0.984 (0.0290)
AIC Criterion	23596.02		23621.72	

Note: All estimates in odd ratios. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Migrant groups: (G1) Origin of migrants: 123 countries + 7 aggrupation of “Other countries”; (G2) Language groups: English, French, Spanish, Portuguese, German, Swedish, Russian, Arabian and Other; (G3) Regions: EU (3 groups), Non-EU Europe, Maghreb, Sub-Saharan Africa, North America, Central America and the Caribbean, South America, East Asia, Western Asia, Southeast Asia and Oceania

Area of residence: (A1) NUTS2: 17 territorial units (*Comunidades Autónomas*); (A2) NUTS3: 50 territorial units (*Provincias*). In both cases, two autonomous cities (Ceuta and Melilla) are also considered

¹³ For the sake of simplicity, we just present estimation results for our variable of interest, that is, network and the control variable (enclave), as the estimation results for the rest of covariates hardly change across alternative definitions of ethnic groups or areas of residence. All the omitted results are available upon request.

Some interesting findings emerge from the new specifications. First, our main result regarding networks holds regardless of the definition of the group of reference and the area of residence: networks are only relevant for the employee/self-employment decision. Second, the impact of networks on the probability of becoming self-employed (first stage) is always positive for the narrowest definition of area of residence (NUTS 3), independently of how we define the group of reference of migrants. In other words, the closer migrants in the same group live, the more likely the network affects the self-employment decision. Third, among the estimates using NUTS 3 areas of residence, the definition of the group of reference of migrants (country of birth -G1-, language group -G2- or world region -G3-) seems to be relevant, as the impact of networks is only significant for the common language networks. However, the AIC indicates that the global equations are preferred for the less broad definitions. Consequently, in the next section we check the relevance of the number of countries included in the estimations.

5.2.2. *Stability across years in the sample*

All the results previously shown are obtained for the pool estimation of our data obtained for the period 1999-2009. To check for the existence of differences in our estimates across time, we split the sample into two periods: 1999-2007 and 2008-2009 and estimate our preferred sequential logit specification using each subsample separately. We select this breaking point because of the fact that in 2008 a change in the structure of our database happened, as the number of migrants and countries of origin included in the sample experienced a relevant quantitative leap (85 countries on average in the 1999-2007 sub-sample versus 104 countries in the 2008-2009 sub-sample).

The estimation results are shown in Table 6. These results hold remarkably faithful to those in Table 3. The main difference can be found for the Transition II in the period 2008-2009, as the effect of educational attainment is not significant; therefore, the years of residence are the only determinant of the entrepreneur option in that case. This outcome, in fact, reinforces our main result about the increasing relevance of the personal experience in the home country for becoming a job creator.

Table 6. Sequential Logit Estimates. Robustness analysis

First Sub-sample: 1999-2007

Second Sub-sample: 2008-2009

VARIABLES	Pool 1999-2007		Pool 2008-2009	
	Transition I	Transition II	Transition I	Transition II
<i>Network_{jk}</i>	1.005*** (0.0007)	0.999 (0.0002)	1.006*** (0.0012)	1.000 (0.0004)
<i>Age_{ijk}</i>	1.077*** (0.0183)	1.076** (0.0353)	1.027* (0.0198)	0.979 (0.0401)
<i>Age²_{ijk}</i>	0.999* (0.0002)	0.999** (0.0004)	1.000 (0.0002)	1.000 (0.0004)
<i>Female_{ijk}</i>	0.469*** (0.0265)	0.833* (0.0899)	0.479*** (0.0329)	1.040 (0.1393)
<i>Married_{ijk}</i>	1.388*** (0.0827)	1.039 (0.1219)	1.443*** (0.1056)	1.112 (0.1739)
<i>Years of residence</i>				
$1 < YR_{ijk} \leq 5$	1.177* (0.1151)	1.232 (0.2830)	1.441** (0.2567)	4.355* (3.3093)
$5 < YR_{ijk} \leq 10$	2.078*** (0.2162)	1.651** (0.3782)	1.921*** (0.333)	9.7651*** (7.4394)
$YR_{ijk} > 10$	2.775*** (0.2990)	1.992*** (0.4574)	3.062*** (0.5394)	13.758*** (10.603)
Educational attainment				
<i>Primary_{ijk}</i>	0.899 (0.1051)	1.762** (0.4878)	1.185 (0.1894)	1.378 (0.4637)
<i>Secondary_{ijk}</i>	1.315** (0.1584)	1.661** (0.4715)	1.493*** (0.2430)	1.291 (0.4399)
<i>Tertiary_{ijk}</i>	1.501*** (0.1817)	1.846** (0.5264)	1.606*** (0.2659)	1.575 (0.5366)
<i>Enclave_{jk}</i>	0.894*** (0.0195)	1.017 (0.0119)	0.864*** (0.0223)	0.995 (0.0307)
<i>Constant</i>	0.014*** (0.0059)	0.039*** (0.0338)	0.006*** (0.0048)	0.00000002*** (0.00000002)
Observations	20,369	20,369	12,269	12,269

Note: All estimates in odd ratios. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Fixed effects: (A) Area of residence: *Spanish NUTS2*. (B) Area of origin: *EU (3 groups), Non-EU Europe, Maghreb, Sub-Saharan Africa, North America, Central America and the Caribbean, South America, East Asia, Western Asia, Southeast Asia and Oceania*. (C) Sectors: *Agriculture, Manufacturing and Services*. (D) Years: 1999 to 2007 in Equations (1)-(2) and 2008-2009 in Equations (3)-(4).

6. CONCLUSIONS

In this paper we have studied the relevance of migrant networks on the decision of foreign workers to become self-employed in the host country, distinguishing, for the first time in the literature, between own-account workers and job creators. Our analysis suggests that distinguishing between the decision making process between own-account workers and entrepreneurs is important because it enables us to highlight the dynamic process of assimilation of migrants. Specifically, empirical tests show the superior performance of a sequential choice logit model against a simultaneous choice model implying that migrants choose first between employee and self-employment options (first transition) and then they face the decision between own-account worker and entrepreneur if they choose self-employment (second transition). In the first transition, the ethnic network, the educational attainment and the number of years of residence within the host economy of the individuals affect positively the probability of self-employment. In the second transition, the existence of networks has no impact; only the individual characteristics affect positively the likelihood of becoming entrepreneur. The results are robust to alternative definitions of geographic areas where migrants live in the host country as well as definitions of groups of migrants (by country of birth, linguistic groups or broad continental regions).

Some important policy implications can be derived from our results. In general, our findings point to a dynamic process of assimilation where different factors influence decisions at different stages of that process. Thus, the extent of welfare-enhancing entrepreneurship stimulated by migration requires different interventions at different times in the process with our results suggesting that entrepreneurship among foreign born population requires a minimum years of residence in the host economy. In particular, networks play a positive role in helping migrants to become self-employed which suggests migrants could benefit from active policies aimed to provide information on the local labour markets and financial resources for starting up new business. Our findings suggest that such information provision would be more important for individuals who lack an active community of co-nationals to support them. Second, individuals who are prepared to become entrepreneurs appear to have acquired both the knowledge and the resources and no longer need assistance from co-nationals.

REFERENCES

- Andersson, L. and Hammarstedt, M. (2012). Ethnic Enclaves, Networks and Self-Employment among Middle Eastern Immigrants in Sweden, *International Migration*, 53 (6), 27-40. doi:10.1111/j.1468-2435.2011.00714.x
- Bartel A. (1989). Where Do the New U.S. Immigrants Live? *Journal of Labor Economics*, 7 (4), 371-391. <http://dx.doi.org/10.1086/298213>
- Blanchflower, D.G. (2000). Self-employment in OECD countries, *Labour Economics*, 7 (5), 471-505. [https://doi.org/10.1016/S0927-5371\(00\)00011-7](https://doi.org/10.1016/S0927-5371(00)00011-7)
- Bertrand, M., Luttmer, E.F.P and Mullainathan, S. (2000). Network Effects and Welfare Cultures, *The Quarterly Journal of Economics*, 115 (3), 1019-1055. <https://doi.org/10.1162/003355300554971>
- Borjas, G.J. (1986). The Self-Employment Experience of Immigrants, *The Journal of Human Resources*, 21 (4), 485-506.
- Borjas, G.J. (1999). Immigration and Welfare Magnets, *Journal of Labor Economics*, 17 (4), 607-637. <https://doi.org/10.1086/209933>
- Buis, M. L. (2010). Chapter 6, Not all transitions are equal: The relationship between inequality of educational opportunities and inequality of educational outcomes", In: Buis, Maarten L. *Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands during the 20th Century*, PhD thesis. http://www.maartenbuis.nl/dissertation/chap_6.pdf
- Buis, M.L. (2011). The Consequences of Unobserved Heterogeneity in a Sequential Logit Model, *Research in Social Stratification and Mobility*, 29 (3), 247-262. <http://dx.doi.org/10.1016/j.rssm.2010.12.006>
- Cameron, S. V., Heckman J. J. (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *The Journal of Political Economy*, 106 (2), 262-333. <http://dx.doi.org/10.1086/250010>

- Carroll, R., Holtz-Eakin, D., Rider, M. and Rosen, H.S. (2000). Income Taxes and Entrepreneur's Use of Labor, *Journal of Labor Economics*, 18 (2), 324-351. <https://doi.org/10.1086/209961>
- Chiswick B.R. and 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
- Clark, K., Drinkwater, S. and Robinson, C. (2017). Self-employment amongst migrant groups: new evidence from England and Wales, *Small Business Economics*, 48 (4), 1047-1069. <https://doi.org/10.1007/s11187-016-9804-z>
- Constant, A. and Zimmermann, K.F. (2006). The Making of Entrepreneurs in Germany: Are Native Men and Immigrant Alike? *Small Business Economics*, 26 (3), 279-300. <https://doi.org/10.1007/s11187-005-3004-6>
- Cowling, M., Taylor, M. and Mitchell, P. (2004). Job Creators, *The Manchester School*, 72 (5), 601-617. doi:10.1111/j.1467-9957.2004.00411.x
- Cueto, B. and Rodríguez Álvarez, V. (2015). Determinants of Immigrant Self-Employment in Spain, *International Journal of Manpower*, 36 (6), 895-911. <https://doi.org/10.1108/IJM-10-2013-0233>
- Earle, J.S. and Savoka, Z. (2000). Business star-ups or disguised unemployment? Evidence on the character of self-employment from transition economies, *Labour Economics*, 7 (5), 575-601. [https://doi.org/10.1016/S0927-5371\(00\)00014-2](https://doi.org/10.1016/S0927-5371(00)00014-2)
- Edin, P., Fredriksson, P. and Åslund, O. (2003). Ethnic enclaves and the economic success of migrants – Evidence from a natural experiment, *The Quarterly Journal of Economics*, 118 (1), 329-357. <https://doi.org/10.1162/00335530360535225>
- Fairlie, R.W. and Meyer, B.D. (1996). Ethnic and Racial Self-Employment Differences and Possible Explanations, *The Journal of Human Resources*, 31 (4), 757-793
- Foreman-Peck, J. and Zhou, P. (2013). The Strength and Persistence of Entrepreneurial Cultures, *Journal of Evolutionary Economics*, 23 (1), 163-187. <https://doi.org/10.1007/s00191-011-0239-z>

- Gindling, T.H. and Newhouse, D. (2014). Self-Employment in the Developing World, *World Development*, 56, 313-331. <https://doi.org/10.1016/j.worlddev.2013.03.003>
- Greene, W. H. (2012). *Econometric Analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.
- Hammarstedt, M. (2004). Self-Employment Among Immigrants in Sweden – An Analysis of Intragroup Differences, *Small Business Economics*, 23 (2), 115-126. <https://doi.org/10.1023/B:SBEJ.0000027664.58874.62>
- Hausman, J. A., and McFadden, D.L. (1984). Specification tests for the multinomial logit model, *Econometrica*, 52 (5), 1219–1240. doi:10.2307/1910997
- INE (2014) Spain in Figures 2014. National Statistics Institute. <http://ec.europa.eu/eurostat/web/ess/-/spain-in-figures-2014>
- Kautonen, T., Down, S. and Minniti, M. (2014). Ageing and entrepreneurial preferences *Small Business Economics*, 42 (3), 579-594. <https://doi.org/10.1007/s11187-013-9489-5>
- Lassmann, A. and Busch, C. (2015). Revisiting native and immigrant entrepreneurial activity, *Small Business Economics*, 45 (4), 841-873. <https://doi.org/10.1007/s11187-015-9665-x>
- Le, A.T. (2000). The Determinants of Immigrant Self-Employment in Australia, *The International Migration Review*, 34 (1), 183-214.
- Light, I. (1972). *Ethnic Enterprise in America: Business and Welfare among Chinese, Japanese, and Blacks*. Berkeley, CA: University of California Press
- Light, I. (1984). Immigrant and ethnic enterprise in North America, *Ethnic and Racial Studies*, 7(2), 195-216. <http://dx.doi.org/10.1080/01419870.1984.9993441>
- McFadden, D.L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.). *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press
- Nagakura, D. and Kobayashi, M. (2009). Testing the Sequential Logit Model Against the Nested Logit Model, *The Japanese Economic Review*, 60 (3), 345-361. doi:10.1111/j.1468-5876.2008.00458.x

- OECD (2011). *International Migration Outlook 2011*. Paris: OECD Publishing.
http://dx.doi.org/10.1787/migr_outlook-2011-en
- OECD/The European Commission (2014). *The Missing Entrepreneurs: Policies for Inclusive Entrepreneurship in Europe*, OECD Publishing. <http://dx.doi.org/10.1787/9789264213593-en>
- Pietrobelli, C., Rabellotti, R. and Aquilina, M. (2004). An empirical study of the determinants of self-employment in developing countries, *Journal of International Development*, 16 (6), 803-820.
doi:10.1002/jid.1117
- Simoës, N., Crespo, N. and Moreira, S.B. (2016). Individual determinants of self-employment entry: What do we really know?, *Journal of Economic Surveys*, 30 (4), 783-806. doi:10.1111/joes.12111
- Toussaint-Comeau, M. (2012). Ethnic Social Networks and Self-Employment of Immigrant Men in the US, *Eastern Economic Journal*, 38 (1), 74-98. <http://dx.doi.org/10.1057/ej.2010.52>
- Tutz, G. (1991). Sequential models in categorical regression, *Computational Statistics & Data Analysis*, 11 (3), 275-295. [https://doi.org/10.1016/0167-9473\(91\)90086-H](https://doi.org/10.1016/0167-9473(91)90086-H)
- Yuengert, A. M (1995). Testing Hypotheses of Immigrant Self-Employment, *The Journal of Human Resources*, 30 (1), 194-204.

Appendix A.1. Results of the multinomial logit specification

In our search for the best econometric model to identify the impact of networks on the migrants' decision on employee, own-account worker or job creator, we analyse first the scenario where immigrants face all the three options simultaneously (multinomial logit model). These estimates are displayed in Table A.1. using employees as reference category.

Table A.1 Multinomial Logit Specification
Reference category: employee

VARIABLES	Own-Account Worker	Entrepreneurs
<i>Network_{jk}</i>	1.005*** (0.0003)	1.005*** (0.0004)
<i>Age_{ijk}</i>	1.047*** (0.0149)	1.086*** (0.0247)
<i>Age²_{ijk}</i>	0.999 (0.0002)	0.999* (0.0003)
<i>Female_{ijk}</i>	0.482*** (0.0233)	0.453*** (0.0329)
<i>Married_{ijk}</i>	1.373*** (0.0705)	1.518*** (0.1232)
<i>Years of residence</i>		
1 < YR _{ijk} ≤ 5	1.195* (0.1105)	1.528** (0.2954)
5 < YR _{ijk} ≤ 10	1.679*** (0.1592)	3.609*** (0.6816)
YR _{ijk} > 10	2.197*** (0.2126)	5.802*** (1.1028)
<i>Educational attainment</i>		
<i>Primary_{ijk}</i>	0.907 (0.0970)	1.348* (0.2462)
<i>Secondary_{ijk}</i>	1.278** (0.1390)	1.825*** (0.3387)
<i>Tertiary_{ijk}</i>	1.370*** (0.1500)	2.178*** (0.4039)
<i>Enclave_{jk}</i>	0.881*** (0.0097)	0.899*** (0.0117)
<i>Constant</i>	0.0166*** (0.0061)	0.001*** (0.007)
Akaike IC	23239.88	
Observations	32,638	

Note: All estimates in odd ratios. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Fixed effects: (1) Area of residence: *Spanish NUTS2*. (2) Area of origin: *EU, Non-EU Europe, Maghreb, Sub-Saharan Africa, North America, Central America and the Caribbean, South America, East Asia, Western Asia, Southeast Asia and Oceania*. (3) Sectors: *Agriculture, Manufacturing and Services*. (4) Years: 1999 to 2009.

We observe that the effect of networks on the probability of choosing one of the self-employment categories is always significant; the odds-ratio estimates are higher than one, which indicate that the better the network relationships among the co-nationals living in the area, the higher the probability of choosing one of the self-employment categories against the employee option. Besides, most of the personal characteristics are significant and positively affect the chances of self-employment. However, the most striking result is that there are no substantial differences in the impact of these variables on the probability of choosing own-account or entrepreneurship with respect to the reference category (employee).

The similarity in the results for both types of self-employment may suggest that treating the three alternatives as independent and concurrent in the same choice is inappropriate (as confirmed by the Hausman test displayed in Table 2 in the main text).

Appendix A.2. A sensitivity test to unobserved heterogeneity in sequential logit estimates

Buis (2011) proposes to estimate the effect of explanatory variables assuming different scenarios concerning the amount of unobserved heterogeneity in the model to find out whether unobserved factors may cause a bias problem in the estimated coefficients. Such unobserved heterogeneity – a weighted sum of all unobserved heterogeneity sources- is assumed standardised and normally distributed, Z , with an impact β_z on each transition probability that is constant over transitions. According to Buis (2011), the effect of the standardised variable Z , β_z equals the standard deviation of an unstandardized random variable v : $\beta_z Z = v$. Thus, the probability for transition τ is:

$$P_{\tau,ijk} = \Lambda(\beta_{0\tau} + \sum_h \beta_{h\tau} X_{hi} + \alpha_{j\tau} + \gamma_{k\tau} + \beta_z Z)$$

with $Z \sim N(0,1)$ and uncorrelated with migrant networks during the first transition of the sequential model. The sensitivity analysis is based on the specification of extreme but still plausible scenarios on the amount of unobserved heterogeneity pushing the model to explore how estimation results are affected. In other words, we aim to investigate how extreme one scenario needs to be before our conclusions change. The scenarios differ according to the amount of unobserved heterogeneity, and, since Z is a standardised variable, they differ according to the impact of the amount of unobserved heterogeneity, β_z . Concerning

the extreme but still plausible scenario for the upper bound for β_z , our hypothesis assumes that the effect of unobserved heterogeneity may be as large as the largest impact of the fixed effects included in our estimation as they control for other (observed) heterogeneity sources. In our estimation we found that the largest impact was, in odds ratio terms, 4.3. Thus a reasonable value will be: $\beta_z = \ln(4.3) = 1.5$. In order to be sure that we include extreme scenarios, the model is estimated where β_z values range from 0 (for the base-line model without unobserved heterogeneity) to approximately twice the given reasonable value, in line with Buis (2011): $\beta_z \in [0,3]$. Table A.2 shows the sensitivity analysis results for our main interest variables.

In Transition 1 we achieve outcomes quite similar to those in our baseline model (the sequential specification in Table 3): odds ratio estimates for network are larger than 1 and significant, enclave estimates are lower than 1 and significant, and the results for the rest of covariates slightly varying in magnitude but not in its sign or significance. These results are robust to different amounts of unobserved heterogeneity effects. In Transition 2, in general, we obtain almost the same results as in the baseline model. Nevertheless, some estimates slightly change when large unobserved heterogeneity effects are considered: the network variable shows a deviation from the baseline model when the effect of unobserved heterogeneity is higher than the extreme but reasonable scenario we assumed, becoming statistically significant, which suggests an increase in the odds of a self-employed worker being an entrepreneur. Similarly, the positive effect of age, being married or living between 1 and 5 years in the area on the probability of becoming an entrepreneur will be significant from the extreme but still reasonable scenario of unobserved heterogeneity effects we have assumed. Finally, only the effect of being a female will become 5% significant from a lower impact of unobserved heterogeneity.

Table A.2 Sensitivity analysis. Unobserved heterogeneity effects.
Sequential Logit Estimates

	$\beta_z = 0$	$\beta_z = 0.5$	$\beta_z = 1$	$\beta_z = 1.5$	$\beta_z = 2$	$\beta_z = 2.5$	$\beta_z = 3$
Transition I							
<i>Network_{ijk}</i>	1.005*** [0.000]	1.005*** [0.000]	1.005*** [0.000]	1.006*** [0.000]	1.006*** [0.000]	1.007*** [0.000]	1.007*** [0.000]
<i>Age_{ijk}</i>	1.057*** [0.000]	1.057*** [0.000]	1.057*** [0.000]	1.059*** [0.000]	1.063*** [0.001]	1.069*** [0.001]	1.077*** [0.001]
<i>Age²_{ijk}</i>	0.999 [0.298]	0.999 [0.369]	0.999 [0.557]	0.999 [0.776]	0.999 [0.946]	1.000 [0.948]	1.000 [0.889]
<i>Female_{ijk}</i>	0.474*** [0.000]	0.462*** [0.000]	0.429*** [0.000]	0.385*** [0.000]	0.336*** [0.000]	0.289*** [0.000]	0.246*** [0.000]
<i>Married_{ijk}</i>	1.409*** [0.000]	1.426*** [0.000]	1.476*** [0.000]	1.553*** [0.000]	1.652*** [0.000]	1.771*** [0.000]	1.908*** [0.000]
<i>Years of residence</i>							
1 < <i>YR_{ijk}</i> ≤ 5	1.248*** [0.009]	1.254*** [0.009]	1.275*** [0.008]	1.308*** [0.008]	1.348*** [0.009]	1.392*** [0.010]	1.441** [0.011]
5 < <i>YR_{ijk}</i> ≤ 10	1.994*** [0.000]	2.031*** [0.000]	2.147*** [0.000]	2.341*** [0.000]	2.604*** [0.000]	2.937*** [0.000]	3.345*** [0.000]
<i>YR_{ijk}</i> > 10	2.779*** [0.000]	2.873*** [0.000]	3.164*** [0.000]	3.662*** [0.000]	4.380*** [0.000]	5.348*** [0.000]	6.621*** [0.000]
<i>Educational attainment</i>							
<i>Primary_{ijk}</i>	1.002 [0.982]	1.005 [0.957]	1.015 [0.887]	1.033 [0.785]	1.055 [0.685]	1.081 [0.605]	1.110 [0.542]
<i>Secondary_{ijk}</i>	1.395*** [0.001]	1.413*** [0.000]	1.465*** [0.000]	1.553*** [0.000]	1.674*** [0.000]	1.822*** [0.000]	1.996*** [0.000]
<i>Tertiary_{ijk}</i>	1.542*** [0.000]	1.570*** [0.000]	1.655*** [0.000]	1.793*** [0.000]	1.981*** [0.000]	2.213*** [0.000]	2.489*** [0.000]
<i>Enclave_{jk}</i>	0.887*** [0.000]	0.884*** [0.000]	0.877*** [0.000]	0.868*** [0.000]	0.856*** [0.000]	0.845*** [0.000]	0.833*** [0.000]
Transition II							
<i>Network_{jk}</i>	0.999 [0.946]	1.000 [0.796]	1.000 [0.344]	1.000 [0.106]	1.001** [0.030]	1.001*** [0.008]	1.001*** [0.002]
<i>Age_{ijk}</i>	1.035 [0.172]	1.037 [0.173]	1.043 [0.149]	1.055* [0.098]	1.071* [0.055]	1.088** [0.031]	1.106** [0.019]
<i>Age²_{ijk}</i>	0.999 [0.124]	0.999 [0.141]	0.999 [0.166]	0.999 [0.165]	0.999 [0.149]	0.999 [0.132]	0.999 [0.119]
<i>Female_{ijk}</i>	0.906 [0.233]	0.874 [0.121]	0.786** [0.012]	0.676*** [0.000]	0.571*** [0.000]	0.482*** [0.000]	0.409*** [0.000]
<i>Married_{ijk}</i>	1.092 [0.341]	1.110 [0.278]	1.168 [0.145]	1.259* [0.051]	1.369** [0.015]	1.491*** [0.004]	1.621*** [0.002]
<i>Years of residence</i>							
1 < <i>YR_{ijk}</i> ≤ 5	1.372 [0.130]	1.403 [0.116]	1.500* [0.083]	1.638* [0.054]	1.796** [0.034]	1.963** [0.022]	2.140** [0.015]
5 < <i>YR_{ijk}</i> ≤ 10	2.161*** [0.000]	2.295*** [0.000]	2.715*** [0.000]	3.397*** [0.000]	4.303*** [0.000]	5.424*** [0.000]	6.790*** [0.000]
<i>YR_{ijk}</i> > 10	2.637*** [0.000]	2.874*** [0.000]	3.649*** [0.000]	5.022*** [0.000]	7.041*** [0.000]	9.822*** [0.000]	13.58*** [0.000]
<i>Educational attainment</i>							
<i>Primary_{□□}</i>	1.593** [0.026]	1.633** [0.024]	1.739** [0.020]	1.875** [0.017]	2.016** [0.014]	2.164** [0.013]	2.329** [0.012]
<i>Secondary_{ijk}</i>	1.517* [0.051]	1.572** [0.042]	1.734** [0.024]	1.976** [0.011]	2.268*** [0.005]	2.606*** [0.003]	2.996*** [0.001]
<i>Tertiary_{ijk}</i>	1.752*** [0.009]	1.838*** [0.006]	2.093*** [0.002]	2.484*** [0.001]	2.971*** [0.000]	3.550*** [0.000]	4.237*** [0.000]
<i>Enclave_{jk}</i>	1.013 [0.149]	1.012 [0.197]	1.010 [0.367]	1.006 [0.653]	1.001 [0.969]	0.995 [0.759]	0.990 [0.549]

Note: All estimates in odd ratios. *p-values* in brackets *** p<0.01, ** p<0.05, * p<0.1.

Appendix A.3. Grouping of origin countries of migrants for robustness analysis

Language groups (G2)

Group 1 English	Ireland, United Kingdom, Gibraltar, Canada, US, Australia and New Zealand
Group 2 French	Belgium, France, Luxembourg and Monaco
Group 3 Spanish	Costa Rica, Cuba, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Puerto Rico, Dominican Republic, Argentina, Bolivia, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay and Venezuela
Group 4 Portuguese	Portugal, Brazil, Cape Verde.
Group 5 Russian	Belarus, Russia, Kazakhstan* and Kyrgyzstan*
Group 6 Arabian	Algeria, Egypt, Libya, Morocco, Occidental Sahara, Tunisia, Mauritania*, Saudi Arabia, Iraq, Jordan, Lebanon, Syria, Palestine*
Group 7 German	Germany, Austria and Liechtenstein.
All other countries not included on a specific language groups are included individually in the regression	

World Regions (G3)

Group 1 EU members	
Group 1A EU members before 1999	Germany, Austria, Belgium, Denmark, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, United Kingdom and Sweden.
Group 1B EU members entering 2004	Cyprus, Slovenia, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Czech Republic and Slovakia.
Group 1C EU members entering 2007	Bulgaria and Romania
Group 2 Non-EU European countries	Albania, Andorra, Armenia, Azerbaijan, Belarus, Bosnia-Herzegovina, Croatia, Georgia, Iceland, Liechtenstein, Macedonia, Moldavia, Monaco, Norway, Russia, Switzerland, Turkey, Ukraine, Gibraltar, Kosovo*, Montenegro* and Serbia*.
Group 3 Maghreb	Algeria, Egypt, Libya, Morocco, Occidental Sahara, Tunisia and Mauritania*
Group 4 Sub-Saharan Africa	Cape Verde, Gambia, Equatorial Guinea, Senegal, South Africa, Sudan, Ethiopia ¹ , Angola ¹ , Democratic Republic of Congo*, Cameroon*, Ghana*, Guinea*, Guinea-Bissau*, Mali*, Nigeria* and Other African Countries*.
Group 5 North America	Canada, United States and Dependent Territories in North America (Bermuda, Greenland, Saint Pierre and Miquelon)
Group 6 Central America and the Caribbean	Costa Rica, Cuba, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Panama, Puerto Rico, Dominican Republic and Other countries and territories in Central America and the Caribbean.
Group 7 South America	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela and Other countries and territories in South America.
Group 8 Far East	China (including Hong-Kong and Macao), North Korea, South Korea, Japan, Mongolia and Taiwan.
Group 9 Middle East	Saudi Arabia, Iraq, Israel, Jordan, Lebanon, Syria, Palestine* and Other countries and territories in Middle East
Group 10 South and South-East Asia	Bangladesh, Cambodia, Philippines, India, Indonesia, Iran, Laos, Pakistan, Sri Lanka, Thailand, Vietnam, Kazakhstan*, Kyrgyzstan*, Tajikistan*, Uzbekistan* and Other countries and territories in South and South-East Asia.
Group 11 Oceania	Australia, New Zealand and Other countries and territories in Oceania

¹ since 2004 in the LFS. In the previous years, they were included in the “Other countries...” categories.

* since 2006 in the LFS. In the previous years, they were included in the “Other countries...” categories.