NOVAFRICA

Nova Africa Center
for Business and Economic Development

Working Paper Series

Assessing the Role of Social Networks on Migrant Labor Market Outcomes

Cátia Batista

Nova School of Business and Economics, CReAM, IZA and NOVAFRICA

Ana Isabel Costa

Nova School of Business and Economics and NOVAFRICA

ISSN 2183-0843

Working Paper No 1601

April 2016

Revised January 2018

Assessing the Role of Social Networks on Migrant Labor Market Outcomes*

Cátia Batista† and Ana Isabel Costa‡

January 2018

Abstract

What role do social networks play in determining migrant labor market outcomes? We examine this research question using a random survey of 1500 immigrants living in Ireland. We empirically test the hypothesis that immigrants with more contacts in the host country perform better in the labor market. Our empirical analysis confirms this prediction by focusing broadly on the relationship between migrants' social networks and a variety of labor market outcomes (namely wages, employment, occupational choice and job security), innovatively relative to the existing literature. We find evidence that having one more close contact person in the host country is associated with an increase of nearly 100 euros in the average monthly net salary, and with a higher probability of having a permanent job contract. Network size also seems to have a positive impact in the probability of migrants entering low-skilled occupations, but no effect on high-skilled occupations. Our data is not strongly supportive of a network size effect on employment. Our results are robust to sample selection and other endogeneity concerns. Overall, this paper expands previous findings in the literature mostly focused on wages and employment, and concludes that networks may also provide job security to immigrants.

JEL classification: D8, F22, J15, J31; J61.
Keywords: International Migration; Social Networks; Wage Determination; Job Search; Labor Market Integration; Occupational Choice; Job Security.

---

* The authors are grateful for helpful suggestions from Bien Balocating, Pedro Brinca, Francesco Cestari, Ana Garcia, Matilde Grácio, Teresa Molina Millán, Ana Balcão Reis, Julia Seither, and José Tavares. This paper makes use of an immigrant survey generously funded by the EU NORFACE Programme on International Migration, under the lead of Principal Investigators Catia Batista and Gaia Narciso.
† Nova School of Business and Economics, CReAM, IZA and NOVAFRICA. Email: catia.batista@novasbe.pt
‡ Nova School of Business and Economics, and NOVAFRICA. Email: costa.anaiml@gmail.com

1. Introduction

The study of social networks emerged in the economics literature as a way to explain information mismatches and other market frictions. Early empirical studies show that about half of employed individuals rely on family and friends to find jobs or to have access to job information (Gregg and Wadsworth, 1996; Addison and Portugal, 2001). The main theoretical hypothesis linking social networks to labor market outcomes proposes the existence of an informal channel, usually formed by relatives, friends and acquaintances that provides individuals information not available through formal sources. Such channel mitigates job search frictions in two distinct ways: by making available information to job seekers about employment opportunities, while also providing employers references about the workers.

In the recent decades, a growing literature has investigated the role of social networks on migrant outcomes. In one of its most prominent findings, migrants tend to cluster in groups and share information among them (Edin et al, 2003; Munshi, 2003), which illustrates the particular importance of social networks for migration. Being newcomers to the labor market, migrants are in need of information about job openings and the characteristics of the labor market, particularly upon arrival to the host country. On the other hand, networks may also be useful from the perspective of the employer who often lacks information about the newly arrived migrants.

As initially proposed by Sjaastad (1962) and Harris and Todaro (1970), migration is often an uncertain and risky investment. However, migration networks can lower the costs and uncertainty of follower migrants, as shown by McKenzie and Rapoport (2010). Umblijs (2012) presents theoretical and empirical evidence that if potential migrants have access to a network at the destination, more risk-averse migrants will migrate than if they had no access to these networks.[1]

Calvo-Armengol and Zenou (2005) were one of the first authors to put forward a theoretical model of the role of networks on labor market outcomes particularly wages and employment. The model accounts for a job-matching process, in which workers find jobs through their social network, and shows that in the steady-state, labor market outcomes are positively correlated across time and across agents within a network. Nonetheless, it assumes that networks are exogenous, a strong assumption that has not been validated by empirical research such as Munshi (2003).

Later, Wahba and Zenou (2005) extended the model of Calvo-Armengol and Zenou (2005) by differentiating between low- and high-educated individuals and comparing the efficiency of using networks to find a job with other search methods. The underlying assumption made by the authors is that low-educated individuals only use informal job search methods to find a job, while high-educated individuals use both formal and informal methods. Considering population density as a proxy for the size of the network in Egypt, they provide empirical evidence that conditional on being employed, the probability of finding a job through friends and relatives, compared to other search methods, is higher in denser areas than in less denser areas.

Given the importance of networks as an informal institution with consequences in the labor market, one would expect its effect to be stronger among migrants (relative to natives) as they are newcomers to the market. Munshi (2003) argues that both migrants and employers at the destination lack full access to information about each other, reason why they are in need of job referrals. Social networks have the role of decreasing the asymmetry of information between both agents. Using a sample of Mexican migrants in the US, the study finds that

[1] For further evidence of the importance of social networks for migration, see Beine et al. (2015), Batista and Umblijs (2014), Batista and Umblijs (2016), and Batista et. al (2017).

migrants with larger networks are more likely to be employed and hold higher paying jobs upon arrival. More recently, Kerr and Mandorff (2015) provide evidence that immigrants in the US tend to cluster in the same occupations as immigrants from the same nationality, corroborating the result by Patel and Vella (2013) that found evidence of a wage premium for those immigrants who choose to work in the most popular occupations of their networks due to possible market power among these groups.

Although the correlation between networks and higher wages may seem intuitive, some other studies have pointed to an opposite direction. For instance, Loury (2006) and Long et. al (2013) find evidence of a wage penalty among those individuals who use networks to find jobs over other search methods. Their view rests on the assumption that workers who do not obtain better job offers through formal channels use networks to find jobs as a last resort. For that reason, network users will hold lower paid jobs compared to non-users.

The literature on the effects of the social contacts in the occupational choice of immigrants is relatively recent and has been finding mixed evidence. For instance, Bentolila, Michelacci and Suarez (2010) find that although social networks may help to find jobs. It can create a mismatch in regarding to the productivity of the workers which leads to wage discounts. On the other hand, Patel and Vella (2013) find a positive earnings effects in those migrants who find jobs in similar positions compared to migrants of the same country.

This paper attempts to shed light on the topic by using a random sample of 1500 immigrants from 110 different nations residing in Ireland in 2010 to understand the impact of migrant social networks on different labor market outcomes. We use detailed survey data on these immigrants and their contacts in the host country, which enables us to measure the social network for each migrant, and to find in this way empirical evidence linking networks to the labor market performance of immigrants. The core hypothesis of this study is that, immigrants with larger networks are more exposed to information and therefore perform better in the labor market of the host country.

Our paper contributes to the existing literature by studying the role of migrant networks on various dimensions of labor market integration. We examine four main questions: (i) Is the probability of being employed higher for individuals with more contacts? (ii) What is the effect of the network size on wages: do immigrants with larger networks have higher wages? (iii) Does the size of migrant networks impacts the job security of immigrants in the labor market?; and (iv) Are migrants with larger networks more prone to enter certain types of occupations?

Identification of the causal effects of network size on migrant outcomes is complicated by the possibility of several endogeneity problems. First, the suspected causal relationship between the migrant network size and the respective labor market outcome can be simultaneously determined by unobservable characteristics. Second, endogeneity can also occur through a selection process that leads some immigrants to choose to participate in the labor market while others do not – which results in a sample selection problem in that it is not possible to observe the wages of those who do not work. In order to tackle these main potential estimation biases, and to ensure the robustness of our results, we estimate several distinct models: Linear Probability Model (LPM), Two-Stage Least Squares (2SLS), Heckit model and IV-Heckit model.

We find evidence that immigrants with a larger network size are more likely to earn higher salaries. Moreover, our empirical analysis indicates that network size influences positively the probability of a migrant having a low-skilled occupation. Our results show that the positive impact of migrant networks on employment is not robust to correcting for endogeneity. We also find that immigrants with larger networks are more likely to have a permanent job contract and less likely to not have any type of contract. This indicates that networks provide more security to the labor market outcomes of immigrants - a new finding to the literature.

The structure of the remainder of the paper is organized as follows. Section 2 describes the methodology used and the identification issues. Section 3 presents the data used and the descriptive statistics. Section 4 includes the empirical results, and finally section 5 concludes.

2. Econometric Framework and Identification Strategy

We aim at identifying the effect of network size on four distinct labor market outcomes of immigrants. The main hypotheses to be tested in the empirical analysis is that migrants with a larger network size are more likely to be employed, to benefit from a higher salary, to choose specific types of occupations, as well as to have more job security in the labor market, defined as a more stable job contract. In this section, we discuss the estimation procedures we will follow to test these hypotheses while addressing the potential endogeneity of the network size.

2.1. Linear Probability Model (LPM) estimation

The relationship between an individual migrant's labor market outcome and her network size can be expressed as:

$$Y_i = \alpha_i + \beta_1 N_i + \beta_2 X_i + \varepsilon_i \quad (7)$$

where Y_i is a binary variable that takes the values 0 or 1; N_i stands for the individual network size; X_i is a set of observable individual characteristics such as age, years of schooling, gender, years in Ireland, marital status or country of origin; and ε_i is the unobservable error term.

To identify our parameters of interest, we initially estimated (7) using a Linear Probability Model (LPM).[2] The main coefficient of interest, β_1, produces consistent estimates conditional

[2] Note that we use the LPM when the outcome variable is the probability of being employed, of having a certain occupation, or of holding a certain type of job contract. For the wage equation, we employ an OLS model, which basically follows the same procedure, differing only in the interpretation of the coefficients.

on the assumption that the error term is not correlated with the network size. However, such assumption is rather strong even after controlling for many observable characteristics. It is likely that unobservable characteristics may simultaneously determine the network size and the labor market outcome of a given individual. For instance, people with more contacts may have a better performance on the labor market not exclusively due to the fact that they have more access to information, but also because their unobserved ability is simultaneously correlated with the number of contacts they have and their professional achievements. The use of an Instrumental Variable (IV) approach may enable us to overcome this endogeneity problem and reestablish the consistency of the results.

2.2.Instrumental Variable (IV) Estimates

The causal effect of the migrant network size on his/her subsequent labor market outcome can be obtained by performing an instrumental variable estimation. The estimation model is in this case given by (8):

$$Y_i = \alpha_i + \beta_1 N_i + \beta_2 X_i + \varepsilon_i$$

$$N_i = \theta_i + \gamma S_i + \vartheta_i \tag{8}$$

where S_i represents the instrument for the network size. Such instrumental variable must be strongly correlated with the network size, but not correlated with the error term.

Instrumental variables are widely used in the empirical literature on social networks to ensure estimation consistency. Several authors have used historical networks, proxied by the stock of immigrants a decade prior to the year of the study, as an IV (see, for example, Patel and Vella 2013; Cortes, 2008). As our data was collected in 2010, we employ the stock of immigrants in Ireland in 2000 for every country in the sample to construct our instrument[3]. In order to introduce individual variation and to correct for possible economic disparities

[3] Data for the stock of migrants was made available by the United Nation, Department of Economic and Social Affairs for the years 1990, 2000, 2010 and 2013.

between countries that may affect migration flows, we compute the instrumented size of the network for each migrant in the sample using the following expression:

$$stock_{\ g,2000} * \frac{GDPpc\ PPP_{g,t}}{GDPpc\ PPP_{Ireland,t}}$$

where $stock$ is the stock of immigrants in Ireland in the year 2000; g stands for the immigrant's country of birth; GDPpc PPP represents the per capita Gross Domestic Product in Purchasing Power Parities; and t corresponds to the year of arrival in Ireland of immigrant i.[4]

To construct the instrument for the network size we use the stock of immigrants from the same country in the year 2000. One of the requirements to be part of the study was to have arrived in Ireland between the year 2000 and six-months prior to the interview. For that reason, we do not expect the stock of immigrants ten years prior to the interview to affect labor market outcomes of the surveyed migrants. Furthermore, the use of the GDP per capita in PPP of the countries of origin compared to the Irish GDP in the migration year, helps correcting for economic differences between countries, which is of importance once our sample includes a total of 110 countries around the world.

We confirm in our data that the size of the individual's network is correlated with the stock of immigrants, regardless of the fact that our main definition of networks includes migrants and non-migrants from a variety of countries, as long as they participate in the labor market. It is important to note that the majority of the individuals in the network are in fact immigrants coming from the same country.

2.3.Heckit and IV-Heckit Models

Although the IV approach is useful to correct for the endogeneity problems described before, it cannot deal with other sources of bias - namely that caused by sample selection.

[4] This approach has also been widely used in the literature. See, for example, McKenzie and Rapaport (2010) and Batista et al. (2012).

Indeed, when the dependent variable is the monthly wage, and given the fact that our sample includes individuals that were not working at the time the survey was conducted, sample selection bias may arise. Once a migrant arrived in the destination country he can choose whether to work or not. For instance, some migrants choose to stay at home with their family, while others are still pursuing their studies. If such decisions were randomly made, selection bias would not be a problem and we could proceed with the OLS and IV estimations. However, unobserved factors may be operating a selection mechanism: the decision of entering the labor market may be correlated with a certain type of omitted characteristics that also influence the performance of the migrant on the labor market. Thus, sample selection bias will lead to inconsistent estimations.

To overcome this problem, we estimate the sample selection correction procedure suggested by Heckman (1979), usually known as Heckit, that allows to not only test for sample selection bias, but also to obtain consistent estimates. Then, we further combine the Heckit and IV approaches to simultaneously deal with the two types of endogeneity described above. In what follows, we present the models used to estimate Heckit and IV-Heckit models.

A sample selection model is specified by two equations: an observation or regression equation and a selection equation. The first equation considers the mechanisms determining wages:

$$w_1 = \alpha_1 N_1 + \beta_1 \boldsymbol{z}_1 + u_1 \; observed \; only \; if \; y_2 = 1 \qquad (9)$$

where w_1 is a latent endogenous variable representing the monthly net wage; N_1 stands for the network size and $\boldsymbol{z}_1$ is the vector of other explanatory variables: age, gender, years in Ireland, years of schooling and marital status.

The second equation, known as the selection equation, considers a proportion of the sample for whom the wage is observed, and the mechanism determining the selection process:

$$y_2 = \begin{cases} 1 \; if \; y_2^* > 0 \\ 0 \; if \; y_2^* \leq 0 \end{cases} \qquad (10)$$

$$y_2^* = \theta\boldsymbol{x} + v_2$$

where $\boldsymbol{x}$ represents the set of explanatory variable determining the decision of working[5]: the household income and if the migrant has a child. Note that the selection equation (10) specifies that wages are only observed for those migrants whose market wage is greater than zero. That is, they are only considered to be working if their wage is above a threshold value. A zero value in the equation means that the market wage of migrants is greater than their reservation wage (w>w*).[6]

For the case of the endogeneity of the main explanatory variable, the IV-Heckit model, we introduce the reduced-form equation for the network size, N_1, where $\boldsymbol{\delta}_3$ represents the instrument[7]:

$$N_1 = \boldsymbol{z\delta}_3 + v_3 \tag{11}$$

Heckman (1979) developed a two-stage procedure that allows one to consistently estimate $\boldsymbol{\alpha_1}$. In the first step, using all observations, i.e., the migrants for whom we observe wages or not, a probit model is estimated representing the decision to work: $(y_{2i} = 1|\boldsymbol{\varepsilon}_i) = \Phi(\boldsymbol{\epsilon}_i x)$. Then, based on the parameter estimated $\hat{\delta}_2$, the inverse Mills ratio for each observation is computed. The second step is to use the selected sample, i.e. working migrants, and fit an OLS regression for wages, including the inverse Mills ratio as an explanatory variable: $y_1 = \boldsymbol{x}_1\boldsymbol{\beta}_1 + \gamma_1\lambda(\boldsymbol{x}\hat{\beta}_2) + \varsigma_i$. By including the Mills ratio, $\lambda\left(\boldsymbol{x}\hat{\beta}_2\right)$, as an additional explanatory variable we correct the sample selection problem. For this reason, if the inverted Mills is statistically significant, we can conclude for selectivity bias.

---

5 We acknowledge the fact that some working migrants did not report their wage. In our methodological approach, we estimated the regressions including and excluding them from the sample and the significance of the main explanatory variable remained unchanged. For that reason, we present our results including the whole sample of immigrants.

6 The Heckit model assumes that: (i) $(\boldsymbol{x}, y_2)$ are always observed; (iii) (u_1, v_2) is independent of $\boldsymbol{x}$ with zero mean; (iii) $u_1 \sim N(0,\sigma)$; (iv) $v_2 \sim N(0{,}1)$; (v) $corr(u_1, v_2) = \rho$; and (vi) u_1 and v_2 are error terms of the two regression equations, assumed to be bivariate and normally distributed.

7 The IV-Heckit model incorporates both the assumptions mentioned before and the assumptions of the instrumental variable approach – valid and relevant instruments.

For the IV-Heckit, in the first step, we estimate a probit model for the likelihood of the migrant working, where we include the exogenous explanatory variables, the exclusion restrictions and the instrument. This will give us estimates for $\hat{\delta}_3$. We then compute the inverse Mills ratio for each observation. In the second step, we use the selected subsample and estimate $y_1 = \boldsymbol{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \gamma_1\hat{\lambda}_{i3} + e_1$ by 2SLS using instruments $(\boldsymbol{z}, \hat{\lambda}_{i3})$.

It is important to introduce exclusion restrictions to distinguish a sample selection from a misspecified model (Wooldridge 2006; Cameron and Trivedi 2005). They should have a high impact on the probability of the migrant entering the labor market but not be included in the outcome equation. Note that if we only had one exclusion restriction, in the case of the IV-Heckit model, predicted y_2 would be nearly collinear with $\boldsymbol{z}_1$ and $\hat{\lambda}_{i3}$. For that reason, we include at least two exclusion restrictions in our estimation models. We use the household income not earned by the respondent and a dummy variable for whether the respondent has a child as the two determinant variables influencing the probability of the migrant entering the labor market.

3. Data: context, description and descriptive statistics

3.1. Data Collection

The database employed in our analysis consists of the baseline survey conducted among 1500 eligible individuals, including detailed information of the respondents and their networks[8], who arrived in Ireland between the year 2000 and six months to the interview date[9]. The baseline was collected in the greater Dublin area, between February 2010 and

[8] For more details regarding the project, see Batista and Narciso (2017).

[9] Other requirements to be fulfilled to be part of the project were: 1) to be 18 years-old or older; 2) not Irish or British born; 3) not an Irish or British citizen. As reported in Batista and Narciso (2017), *"British citizens were excluded due to the close historical ties between Ireland and Great Britain."*

December 2011, as part of a larger experimental project on migration and information flows led by a team of researchers at Trinity College Dublin.

Survey activities were conducted by Amarach Research, a reputable survey company with experience conducting research surveys in Ireland, under the close supervision of the authors and their research team.

The random sampling procedure followed three steps: first, 100 Enumeration Areas (EAs) were randomly selected out of the 323 Electoral Districts. This selection was performed according to probability-proportional-to-size sampling, using the 2006 Census of Ireland to define the total number of non-Irish and not-British individuals residing in Ireland. Second, 15 households were chosen within each EA using a random route approach[10]. Finally, if the household had more than one member eligible to take part in the survey, the individual respondent was randomly selected based on a next-birthday rule[11]. Due to missing relevant information about eligibility for nine respondents, the final number of immigrants included in the sample is 1,491.

The random route approach consisted of the following procedure: each enumerator was given a map of the assigned EA and a pre-selected random starting address within the allocated EA; after a successful interview, enumerators were instructed to exit the house, turn left, count five houses down and approach this new address[12]; in the case of an absent household, interviewers were requested to call back to the address for a maximum of five times, at different times of the day and different days of the week. Each call-back was recorded on the interviewer's report. When an address was exhausted after five call-backs, or deemed ineligible, or in the case of a refusal, the interviewer followed predefined instructions

[10] The 15 households are drawn from the non-Irish/non-British population.

[11] The next-birthday rule consists in selecting the person from the household who had the next birthday among all the members of the household aged 18 or older. This procedure is used to randomize the target respondent within the household.

[12] A set of standard rules were given in the case of cross-roads, apartment buildings, and cul de sac.

in order to get the next address, namely the address next door to the left when exiting the house.

All enumerators were initially trained by the research team and were subsequently supervised by the survey company and, randomly, by members of the research team. Each enumerator had to complete an enumeration report, listing each address approached, the number of call-backs and the outcome of each visit. The enumeration reports were closely inspected and verified by the research team. If the randomization instructions were not followed, interviews had to be replaced.

3.2. Defining Network Size

In order to estimate the effect of the network size on wages in the context of migration one first needs to define the network size for each immigrant. We follow (Patel and Vella, 2013) and include in the network the individuals that were working at the time the survey was conducted, and therefore were more able to provide inside information about the labor market. We also tested network definitions including individuals that were not working. However, we obtained weaker results, which suggests that labor market information is shared by those who are insiders in the labor market. This suggests that the quality of the network is important to produce better achievements in the labor market (Calvo-Armengol and Jackson, 2004).

We use three different questions in the survey to cover all the contacts of the immigrants in Ireland. The first one concerns the composition of the household members of the respondent[13]: "Please indicate all the persons who belong to this household". The second question concerns the contacts who the respondent knew living in Ireland before he had moved to the country: "Before coming to Ireland, how many people did you know who were already living in Ireland (at the time you moved)"?[14]; The third question includes the contacts

[13] Household members are defined as those who usually sleep and eat in the same unit.
[14] People belonging to the same household are excluded from this question. We included these contacts since they may act as an important information channel for migrants both before and after the migration process.

living in Ireland whom, at the time of the survey, the migrant had most contact with: “Currently, who are the people (excluding people you live with or people you knew before coming to Ireland) that you have most contact with in Ireland?”. Using this information we are able to observe the size of the representative network for each migrant.

3.3. Descriptive Statistics

Table 1 includes the main descriptive statistics about the migrants’ characteristics. The first variable corresponds to the current monthly net wages expressed in euros. As we can observe, only the mean value of the monthly wage is 1185 euros. As we use a representative migrant sample, part of respondents did not reported a positive wage[15]. Our data indicates that around 75% of the individuals stated to be working as their main occupation. The size of the network is the explanatory variable of main interest. Network size is a discrete variable that assumes values between 0 and 13 in our data. The average number of people in the network is approximately three. The sample is made up of a highly educated class of migrants who had, on average, 15 years of education, which roughly corresponds to a Bachelor’s degree. The average age is approximately 33 years old. The average number of immigrants per country is 10004. Finally, the average monthly household income, excluding the respondent’s, is 1128 euros and around 46.5% of the immigrants have a child.

Tables 2 and 3 show the distribution of foreign-born individuals by continent and top-nationalities. A large percent of the sample (43.93%) is represented by immigrants coming from Africa followed by Europe (32.39%). The tree top nationalities are Nigeria, Poland and India with 19.52%, 10.87% and 6.10%, respectively.

[15] Those individuals for whom we did not observe a positive wage have the following occupations: unpaid housework, student, retired, unemployed, or not allowed to work due to visa issues.

4. Estimation Results and Discussion

4.1. Employment

We previously stated that one of our main hypothesis predicts that larger networks are positively correlated to the likelihood of being employed. Table 4 presents the empirical analysis of this prediction. *Employed* is a binary variable that assumes value 1 if the individual is employed and value 0 if otherwise[16]. We start by using the LPM to study the relationship between the employment status and our main variable of interest, network size.

The results indicate that having one more person in the network is associated with an increase of 2.78 percentage points (pp) in the probability of being employed – an estimate significant at the 1% significance level. Moreover, females are less likely to be employed compared to males, as the gender coefficient is negative and highly significant. Being more educated is also positively correlated with employment, one more year of school increases the probability of being employed by 1.27 pp. Age and age squared are also strong determinants of the probability of being employed, although inversely related. As the immigrant gets older, he is less likely to be employed. Interestingly, the number of years since migration does not seem to influence the employment. The same holds true for marital status.

Although the results are significant they are not robust once endogeneity concerns are accounted for. The network size effect becomes higher in magnitude but non-significant in equation (2). Our instrumental variable seems to be valid as it passes the weak identification test: the Cragg-Donald Wald F test presented in the end of the table is 19.769 implying a strong association between the stock of immigrants and the number of contacts in the network size.[17] Equations (3) and (4) do not account for continent fixed effects and clustered standard errors. They imply the same conclusion that the network size is significant in the linear-probability model but loses significance after the endogeneity correction.

[16] We consider as employed individuals who reported to be working as the main occupation.
[17] Staiger and Stock (1997) indicate that, for the case of a single endogenous regressor, the instrument is strong if it passes the threshold of 10 in the first-stage F-statistic.

4.2. Wages

To empirically test the hypothesis that individuals' wages are a positive function of the size of the informal network, we employ three different models: OLS, Heckit and IV-Heckit.

The estimated coefficients and respective standard errors are presented in Table 5. Monthly wage appears in the logarithmic form given the skewness present in the variable (1.3612, a considerable right-hand skewness).

We begin our analysis by introducing the OLS estimation results (Column 1). Network size is statistically significant at the 5% level with a coefficient of 0.0342. That is, having one more person in the network size is associated with a 3.42 percent increase in wages. Individual's gender does not seem to be correlated with the salary earned by the migrant. The coefficient is negative (suggesting that, on average, females tend to receive less than males), although it is not significant. Moreover, the number of years of schooling is positively correlated with the wages earned. One more year in school is associated with a 2.14% increase in wages, a result significant at the 10% level. The number of years in Ireland since migration, is also an important positive determinant of wages, indicative of the presence of a process of acquisition of human capital in the host country. Married immigrants do not differ on average compared with other marital status. Finally, age and age squared are not correlated with wages in the OLS regression. In addition to the main individual characteristics explaining wages, namely gender, years of schooling, years in Ireland, marital status, age and continent fixed effects, we include further controls that may also potentially affecting wages, to increase the comparability of the different estimation models we use.

Although the OLS estimator allows one to have a first insight on the relation between network size and wages, it may yield biased estimates as sample selection and other endogeneity concerns might be active in our sample. To test and correct for the hypothesis of sample selection, we employ a Heckman selection model, included in the second column of

Table 5. The inverse Mills ratio representing the latent selection factor is negative and statistically significant, indicating that having a child and a higher household income is negatively correlated with the probability of entering the labor market. We can conclude that there is a selection bias on unobserved characteristics that turns the sample of those migrants for whom we observe wages different from the remaining ones. Considering that unobserved characteristics of the immigrants reflect their unobserved ability, the Heckit estimate of network effects reveals that column (1) overestimated the network effects of less able people who gain more from the informal channels. A brief reflection on why less able people take more advantage from their networks leads us to consider that lower earning-ability migrants may be relying more on their social network as they find it more difficult to acquire jobs through formal methods.

Comparing the OLS with the Heckit estimates, we can see that when we take into account selection in the decision of working on unobserved ability, the network effect is still significant at the 10% level and our main coefficient of interest is now 0.0293, lower than the OLS estimate. After correcting for sample selection, age and age squared become significant in opposite directions. While age is positively related with the migrants' wages, age squared influences wages in a negative way. This implies a diminishing marginal effect of age, i.e., as the migrant gets older, the effect of age on wages lessens. This result is robust to the IV-Heckit model. All the other coefficient estimates remain similar to the OLS model.

Column 3 presents the results of the Heckit model with the network size instrumented, our IV-Heckit model. After controlling for endogeneity, we obtain our best estimate of the causal effect of networks on wages, which is still positive and significant at the 5% level. Having one more person in the network is associated with an increase of 8.29 percent in wages. This result suggests that both the OLS and Heckit model underestimate the effects of network size. Our instrumental variable is significant and valid with a first-stage F value of 28.887.

The IV-Heckit results are more robust when compared to the other two models, as it simultaneously deals with the two main sources of potential estimation biases in our empirical analysis. Nevertheless, the robustness of our main coefficient of interest in the three models provides clear evidence that the size of migrant social networks has a positive impact on labor market outcomes. Table 6 presents the first-stage regressions of the Heckit and IV-Heckit models.

4.3. Occupational Choice

In addition to employment and wages, we also examine the impact of network size on other labor market outcomes, namely occupational choice.

As proposed by Patel and Vella (2013) and Kerr and Mandorff (2015), immigrants tend to choose the same occupations as immigrants from the same background, and therefore migrants belonging to the same network may enjoy a large market power in some given sectors of the economy. In our empirical analysis we test if having more contacts can influence individuals to choose a certain type of occupation.

Tables 7 to 9 include the estimations of the immigrants' main occupations. The first table presents the aggregated estimations for the low-skilled labor positions: Agriculture, Industry and Construction. We find mixed evidence of whether having more contacts influences the likelihood of having these type of occupations. The results are significant if we do not account for continent fixed effects in equation (2) and after correcting for endogeneity concerns in equations (3) and (4). The positive coefficient indicates that having more contacts is associated with a higher likelihood of holding a job in one of the three sectors. Estimations in equation (3) show that having one more contact in the network is associated with a 5.75 pp increase in the probability of choosing one of these type of occupations.

Interestingly, evidence is smaller for more high-skilled positions such as those included in the Health and IT sectors as presented in tables 8 and 9, respectively. The negative coefficient of the equations indicates that a larger network size may be negatively associated with the likelihood of entering these types of jobs suggesting a possibility of mismatching through social networks (Bentolila et al. 2010).

4.4. Job security

Although job security seems to have an important role in the labor market performance of immigrants, it has been largely absent from the economics literature on immigrant integration. Job security, i.e., the employment stability provided by the type of labor contracts offered to immigrants may be particularly important for immigrants given their potential vulnerability in host economies. The impact of social networks on this labor market outcome can hence be greatly relevant for migration decisions and outcomes.

We examine the probability of having a certain type of job contract when the network size is larger. Table 10 presents the results for those individuals who have a permanent contract and also for those who have no contract. The results are robust and significant. The network size has a positive impact in the probability of having a permanent job contract. One more contact is associated with a 13.5 pp increase in the probability of having a more secure job contract. We also find that a larger network is associated with a lower probability of holding a job without a contract. Equation (3) and (4) show the LPM and IV estimates. The main coefficients of interest are negative and significant. After correcting for endogeneity, we identify that having one more contact in the network is associated with a lower likelihood of 14.7 pp of having no job contract.

Table 11 presents the same estimations for the probability of having a temporary contract and being self-employed. We do not find evidence of a correlation between migrants holding these types of contracts and the number of contacts in their social networks.

We conclude that social networks seem to provide immigrants with more stable job contracts, and therefore we conclude that one of our main hypotheses is verified empirically.

5. Concluding Remarks

This paper examined the impact of larger network sizes on immigrant's wages, probability of being employed, occupational choice and job security. We find evidence that having more contacts in the host country's labor market is associated with a better performance in the labor market. Having one more contact is associated with an increase of about 98 euros in the average salary. Larger networks also seem to have a higher impact on the immigrants' decision to hold low-skilled occupations when compared with high-skilled occupations and to provide more permanent job contracts and therefore offering more job security. We find weak evidence of an impact on employment.

Expanding the previous findings in the literature that have been mainly focused on wages, and employment, this study concludes that networks may also work as an important way to provide job security to migrants.

Our results may reflect demand and supply influences. On the demand side, employers wanting to reduce their screening costs will employ immigrants based on referrals or, if they are satisfied with the performance of a certain type of workers, they are more likely to hire people from the same background. From the supply perspective, upon arriving in the host country, most immigrants are still seeking for job opportunities (as our data suggests, only 13.48% of them had already a job offer when they moved to Ireland). If individuals are in contact with more people who are employed, it is likely that they will have better information about opportunities in the job market. Although we are not able to disentangle such possible mechanisms in our paper, they provide an interesting topic for future research.

Overall, information flows seem to play an important role for immigrants in a new labor market, a central conclusion for policy makers who wish to create better migration experiences. For that reason, it is important to create platforms in which immigrants can share information about job opportunities in the labor market and therefore benefit other immigrants.

References

Addison John T., and Pedro Portugal. 2002. “Job Search Methods and Outcomes,” *Oxford Economic Papers,* vol. 54: pp. 505-533.

Batista, Catia, and Gaia Narciso. 2017. “Migrant Remittances and Information Flows: Evidence from a Field Experiment”, *World Bank Economic Review*, Forthcoming.

Batista, Catia, and Janis Umblijs. 2014. “Migration, risk attitudes, and entrepreneurship: Evidence from a representative immigrant survey”. *IZA Journal of Migration*, vol. 3: pp. 1-25.

Batista, Catia, and Janis Umblijs. 2016. “Do Migrants Send Remittances as a Way of Insurance?”. *Oxford Economic Papers,* vol. 68 (1): pp. 108-130.

Batista, Catia, Aitor Lacuesta, and Pedro Vicente. 2012. “Testing the ‘Brain Gain’ Hypothesis: Micro Evidence from Cape Verde”. *Journal of Development Economics*, vol. 97(1): pp. 32–45.

Batista, Catia, Tara McIndoe-Calder, and Pedro C. Vicente. 2017. “Return Migration, Self-Selection and Entrepreneurship in Mozambique”. *Oxford Bulletin of Economics and Statistics*, vol. 79: pp. 797–821.

Batista, Catia, Julia Seither, and Pedro C. Vicente. 2017. “Migration, Institutions and Social Networks in Mozambique”, *NOVAFRICA Working Paper* 1701.

Beine, Michel, Frédéric Docquier and Çağlar Özden. 2015. “Dissecting Network Externalities in International Migration”, *Journal of Demographic Economics, vol. 81: 379-408*.

Bentolila Samuel, Claudio Michelacci and Javier Suarez. 2010. “Social Contacts and Occupational Choice”, Economica, vol. 77 (305): pp. 20-45

Calvo-Armengol, Antoni, and Matthew O. Jackson. 2004. The Effects of Social Networks on Employment and Inequality. American Economic Review 94(3), 426–454.

Calvo-Armengol, Antoni, and Yves Zenou. 2005. “Job Matching, Social Network and Word-of-Mouth Communication”, *Journal of Urban Economics*, vol. 57: pp. 500 – 522.

Cortes, Patricia. 2008. "The Effects of Low Skilled Immigration on US prices: evidence from CPI data", *Journal of Political Economy*, vol. 116, n:3.

Edin, P., P.Fredriksson, and P.Aslund. 2003. "Ethnic Enclaves and the Economic Success of Immigrants - Evidence From a Natural Experiment," *Quarterly Journal of Economics*, vol. 118, No. 1, pp. 329—357.

Gregg, Paul, and Jonathan Wadsworth. 1996. “How Effective are State Employment Agencies? Jobcentre Use and Job Matching in Britain”, *Oxford Bulletin of Economics and Statistics,* vol. 58: 443-457.

Harris, John, and Michael Todaro. 1970. “Migration, Unemployment and Development: A two-sector analysis”, *American Economic Review,* vol. 60: 126–42.

Heckman, James. 1979. “Sample Selection Bias as a Specification Error”, *Econometrica,* vol. 47: 153-161.

Kerr, William, and Martin Mandorff. 2015. “Social Networks, Ethnicity, and Entrepreneurship”, *National Bureau of Economic Research Working Paper* 21597.

Long, Wenjin, Simon Appleton, and Lina Song. 2013. “Job Contact Networks and Wages of Rural-Urban Migrants in China”, *IZA Discussion Paper* 7577.

Loury, Linda Datcher. 2006. “Some Contacts Are More Equal than Others: Informal Networks, Job Tenure, and Wages”, *Journal of Labour Economics,* vol. 24, 299-318.

McKenzie, David, and Hillel Rapoport. 2010. "Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks," *Review of Economics and Statistics,* vol. 92(4):811-821.

Munshi, Kaivan. 2003. “Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market,” *Quarterly Journal of Economics,* vol. 118(2), pages 549-5999.

Patel, Krishna, and Francis Vella. 2013. “Immigrant Networks and Their Implications for Occupational Choice and Wages”, *The Review of Economic and Statistics,* Volume 95, Issue 4.

Sjaastad, Larry. 1962. “The Costs and Returns of Human Migration”, *Journal of Political Economy,* vol. 70: pp. 80.

Staiger, Douglas, and James H. Stock. 1997. “Instrumental Variables Regression with Weak Instruments,” *Econometrica,* vol. 65: pp. 557–86.

Umblijs, Janis. 2012. “The Effect of Networks and Risk Attitudes on the Dynamics of Migration”, *Oxford IMI Working Papers*, 54/2012.

Wahba, Jackline and Yves Zenou. 2005. “Density, Social Networks and Job Search Methods: Theory and Application to Egypt”, *Journal of Development Economics,* volume 78, Issue 2, pp. 443-473.

Tables

Table 1. Selected Descriptive Statistics

Variables	Obs.	Mean	Std. Dev.	Max.	Min.
	Main Dependent Variables				
Monthly net wage (euros)	1352	1185.1	1120.7	10500	0
Employed (dummy)	1491	0.7519	0.4321	1	0
	Main Independent Variables				
Network Size	1491	2.870	2.391	13	0
Female (dummy)	1491	0.541	0.498	1	0
Years of schooling	1483	14.69	2.798	17	0
Years in Ireland	1489	5.348	2.863	11	0
Married (dummy)	1491	0.424	0.494	1	0
Age	1491	32.59	8.025	72	18
	Instrumental Variable				
Stock of immigrants in Ireland (year 2000)	1485	10004	12919	44633	2
	Exclusion Restrictions				
Monthly Household Income (euros)	1077	1128	1746	17500	0
Having children (dummy)	1491	0.465	0.499	1	0

Table 2: Distribution of foreign-born individuals in the sample by continent

Continent	Frequency	Percentage
Africa	655	43.93%
Europe	483	32.39%
Asia	247	16.57%
South America	72	4.83%
North America	26	1.74%
Oceania	8	0.54%
Total	**1491**	**100%**

Table 3: Distribution of top-nationalities in the sample

Country of origin	Frequency	Percentage
Nigeria	291	19.52%
Poland	162	10.87%
India	91	6.10%
South Africa	72	4.83%
Romania	63	4.23%
Brazil	54	3.62%
Phillipines	46	3.09%
Total number of countries		**110**

Table 4: Employed as the binary dependent variable

	LPM (1)	IV (2)	LPM (3)	IV (4)
Network Size	0.0278***	0.0316	0.0297***	0.0387
	(5.63)	(0.71)	(6.32)	(1.73)
Female	-0.0652**	-0.0632**	-0.0644**	-0.0639**
	(-2.89)	(-2.80)	(-2.92)	(-2.88)
Years of Schooling	0.0127**	0.0129**	0.0119**	0.0123**
	(2.74)	(2.67)	(3.01)	(3.10)
Years in Ireland	-0.0024	-0.00254	-0.00362	-0.00401
	(-0.51)	(-0.46)	(-0.88)	(-0.96)
Age	0.0402***	0.0400***	0.0405***	0.0396***
	(4.63)	(4.32)	(5.04)	(4.82)
Age^2	-0.0005***	-0.0005***	-0.0005***	-0.0005***
	(-4.11)	(-3.81)	(-4.47)	(-4.22)
Married	-0.0423	-0.0423	-0.0379	-0.0449
	(-1.90)	(-1.02)	(-1.54)	(-1.30)
Constant	-0.203	-0.218	-0.212	-0.225
	(-1.22)	(-1.28)	(-1.49)	(-1.57)
Cragg-Donald Wald F statistic		19.769		67.731
Continent Fixed Effects	YES	YES	NO	NO
N	1481	1475	1481	1475

Standard errors clustered at the Enumeration Area level in equations (1) and (2). *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Log monthly wage as the dependent variable

	OLS (1)	Heckit (2)	IV-Heckit (3)
Network Size	0.0342**	0.0293*	0.0829**
	(3.18)	(2.34)	(2.63)
Female	-0.0398	-0.0181	-0.00571
	(-0.76)	(-0.32)	(-0.10)
Years of Schooling	0.0293*	0.0177	0.0251*
	(2.34)	(1.46)	(2.12)
Years in Ireland	0.0214*	0.0223*	0.0183
	(2.05)	(2.01)	(1.65)
Married	-0.00944	0.0745	0.0735
	(-0.19)	(1.20)	(1.22)
Age	0.0278	0.0757**	0.0754**
	(1.01)	(2.99)	(3.01)
Age^2	-0.000393	-0.00109**	-0.00108**
	(-0.98)	(-3.19)	(-3.22)
Constant	6.056***	5.468***	5.131***
	(15.03)	(11.76)	(11.05)
Continent Fixed Effects	YES	YES	YES
Other individuals characteristics	YES	YES	YES
Labor Market Controls	YES	YES	YES
Inverse Mills Ratio		-0.648**	-0.720***
		(-3.17)	(-3.43)
Cragg-Donald Wald F statistic			28.887
N	900	803	625

Standard errors clustered at the Enumeration Area level in equation (1). *** p<0.01, ** p<0.05, * p<0.1. Other individual characteristics include religion and motive to migrate dummies. Labor market controls include the type of job contract and the first wage in Ireland.

Table 6: Heckman and IV-Heckman's first stage regressions

	Heckit	**IV-Heckit**
	(1)	**(2)**
Network Size	0.133***	
	(3.29)	
Female	-0.157	-0.263
	(-0.89)	(-1.30)
Years of schooling	0.0870**	0.0429
	(2.82)	(1.20)
Years in Ireland	0.00498	0.00949
	(0.16)	(0.27)
Married	-0.190	-0.0599
	(-0.93)	(-0.25)
Age	0.0706	0.177*
	(0.99)	(2.00)
Age^2	-0.000589	-0.00181
	(-0.63)	(-1.54)
Instrumental Variable		-0.00000700
		(-0.78)
Constant		-0.527
		(-0.34)
Continent Fixed Effects	YES	YES
Other individuals characteristics	YES	YES
Labor Market Controls	YES	YES
Continent Fixed Effects	YES	YES
N	803	796

Standard errors clustered at the Enumeration Area in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Agriculture, Industry and Construction sectors as the binary dependent variable

	LPM (1)	LPM (2)	IV (3)	IV (4)
Network Size	0.00461	0.00919**	0.0575*	0.0697***
	(1.48)	(2.77)	(2.15)	(4.69)
Individual Characteristics	YES	YES	YES	YES
Continent Fixed Effects	YES	NO	YES	NO
Constant	0.150	0.130	0.0959	0.0859
	(1.98)	(1.71)	(1.02)	(0.90)
Cragg-Donald Wald F statistic			19.741	67.060
N	1473	1473	1467	1467

Standard errors clustered at the Enumeration Area level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Individual characteristics include gender dummy variable, years of schooling, years in Ireland, age, age squared and married dummy variable.

Table 8: Health sector occupation as the binary dependent variable

	LPM (1)	LPM (2)	IV (3)	IV (4)
Network Size	0.00506	-0.000388	0.0255	-0.0439*
	(1.13)	(-0.09)	(0.80)	(-2.46)
Individual Characteristics	YES	YES	YES	YES
Continent Fixed Effects	YES	NO	YES	NO
Constant	-0.493***	-0.480***	-0.515***	-0.448***
	(-4.95)	(-4.53)	(-4.60)	(-3.92)
Cragg-Donald Wald F statistic			19.741	67.060
N	1473	1473	1467	1467

Standard errors clustered at the Enumeration Area level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Individual characteristics include gender dummy variable, years of schooling, years in Ireland, age, age squared and married dummy variable.

Table 9: IT sector as the binary dependent variable

	LPM (1)	LPM (2)	IV (3)	IV (4)
Network Size	0.00206	0.00366	-0.0483*	-0.00629
	(0.61)	(1.16)	(-2.05)	(-0.57)
Individual Characteristics	YES	YES	YES	YES
Continent Fixed Effects	YES	NO	YES	NO
Constant	-0.0911	-0.0972	0.0398	-0.0903
	(-1.68)	(-1.76)	(-0.48)	(-1.28)
Cragg-Donald Wald F statistic			19.741	67.060
N	1473	1473	1467	1467

Standard errors clustered at the Enumeration Area level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Individual characteristics include gender dummy variable, years of schooling, years in Ireland, age, age squared and married dummy variable.

Table 10: Permanent contract and no contract as the binary dependent variables

	Permanent		**No contract**	
	LMP	**IV**	**LPM**	**IV**
	(1)	**(2)**	**(3)**	**(4)**
Network Size	0.0229**	0.135*	-0.0193**	-0.147*
	(3.11)	(2.16)	(-3.09)	(-2.35)
Female	0.0164	0.00534	0.0441	0.0579
	(0.62)	(0.18)	(1.63)	(1.90)
Years in school	0.0261***	0.0221***	-0.0182***	-0.0141*
	(5.34)	(3.70)	(-3.80)	(-2.35)
Years in Ireland	0.0181**	0.00934	0.00176	0.0115
	(3.06)	(1.28)	(0.31)	(1.57)
Age	0.0239*	0.0159	-0.0364***	-0.0267*
	(2.55)	(1.28)	(-3.98)	(-2.13)
Age^2	-0.000266*	-0.000167	0.000429***	0.000307
	(-2.08)	(-1.00)	(3.52)	(1.83)
Married	0.0923**	0.0101	0.0354	0.130*
	(2.90)	(0.18)	(1.29)	(2.25)
Constant	-0.553***	-0.693**	1.250***	1.404***
	(-3.44)	(-3.20)	(7.04)	(6.46)
Cragg-Donald Wald F statistic		13.829		13.829
N	1271	1267	1271	1267

Standard errors clustered at the Enumeration Area level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Temporary contract and self-employed as the binary dependent variables

	Temporary		Self-employed	
	LMP	**IV**	**LPM**	**IV**
	(1)	**(2)**	**(3)**	**(4)**
Network Size	-0.00858	0.0219	0.00498	-0.00977
	(-1.47)	(0.43)	(1.93)	(-0.48)
Female	-0.0258	-0.0301	-0.0348**	-0.0332***
	(-1.01)	(-1.22)	(-3.36)	(-3.33)
Years in school	-0.0114*	-0.0122*	0.00359*	0.00412*
	(-2.60)	(-2.51)	(2.37)	(2.10)
Years in Ireland	-0.0268***	-0.0289***	0.00691**	0.00805***
	(-5.14)	(-4.88)	(3.20)	(3.37)
Age	0.0125	0.00971	-0.0000178	0.00105
	(1.46)	(0.96)	(-0.01)	(0.26)
Age^2	-0.000172	-0.000136	0.00000948	-0.00000374
	(-1.44)	(-1.00)	(0.22)	(-0.07)
Married	-0.137***	-0.161***	0.00953	0.0206
	(-4.94)	(-3.44)	(1.01)	(1.09)
Constant	0.375*	0.343	-0.0714	-0.0538
	(2.27)	(1.95)	(-1.61)	(-0.76)
Cragg-Donald Wald F statistic		13.829		13.829
N	1271	1267	1271	1267