Language Proficiency, Immigration and Labour Market in Spain

Lisa Abdelhamid

Msc in Economics, Universitat de Barcelona

Advisor: Antonio Di Paolo

Universitat de Barcelona, AQR-IREA, XREAP

July 4, 2018

Abstract

This paper explores the impact of the Spanish proficiency of immigrants on their labour market success in Spain. Using data from the 2007 National Immigrant Survey of Spain, we analyse the effect of language proficiency on several labour market outcomes: employment, contract duration, hours of work, occupational category, self-employment. Applying Instrumental Variables (IV), the results suggest that Spanish proficiency increases the probability of being employed by 16.5pp. This impact is even more relevant for low-educated individuals reaching 31.7pp. Overall, immigrants experienced an improvement in their labour market performance in terms of employment and number of hours. We also found that Spanish proficiency reduces the probability of being self-employed.

Keywords: Immigration, language proficiency, labour market, Instrumental Variable, Spain

JEL Codes: C26, J24, J61, Z13

Acknowledgements

First, I would like to thank my Master thesis advisor Antonio Di Paolo of the Department of Econometrics at the University of Barcelona. With the help of his knowledge and patience, he was able to guide me through every step of this Master thesis.

I would also like to acknowledge the Master program in Economics at the University of Barcelona. For the past two years, the academic members teaching and managing the Master program helped me acquire the required tools to make this thesis possible. I am also thankful to my classmates for the good working atmosphere together at the library.

Finally, I am extremely grateful to my parents for supporting and encouraging me throughout all those years of study.

Contents

"Every immigrant who comes here should be required within five years to learn English or leave the country"

(Theodore Roosevelt in a letter written to the President of the American Defense Society on January 3, 1919)

1 Introduction

The words written by US President, Theodore Roosevelt, at the beginning of the 20th century connect with the many opinions expressed about immigration today. A hundred years after this declaration, in many parts of the world, concerns about the integration of immigrants are part of our daily news and important components of social and economic policies. More specifically, the recent waves of migration to Europe are a constant topic of debate as they represent a large, complex issue for the continent's governments, which struggle to find an appropriate response and seem incapable of implementing a "coherent, long-term and comprehensive strategy" (Metcalfe-Hough, 2015). A significant part of these debates are centred on the systems of integration that countries should implement to promote values of acceptance and tolerance within the shared land. Here, as Roosevelt believed, the host language can play an important role in the integration process of the immigrant. Indeed, the question of language proficiency is worthy of further analysis in the belief that it can give rise to interesting insights for developing immigration policies, not least because labour market assimilation has become a general concern since the non-native born population tend to be more severely affected by unemployment than the native-born (Juchno & Agafitei, 2017).

Many interrelated factors are no doubt responsible for the social gap that emerges between locals and immigrants. The latter have to deal with such obstacles as language barriers, interminable procedures to have their qualifications recognised and a limited network to help in their job hunting efforts. Notwithstanding the obvious challenge considering all the barriers when undertaking a complete integration analysis, here we opt to focus on the relationship between the language proficiency of immigrants and their performance in the labour market. More specifically, we examined the probabilities of employment, contract duration, hours of work and occupational

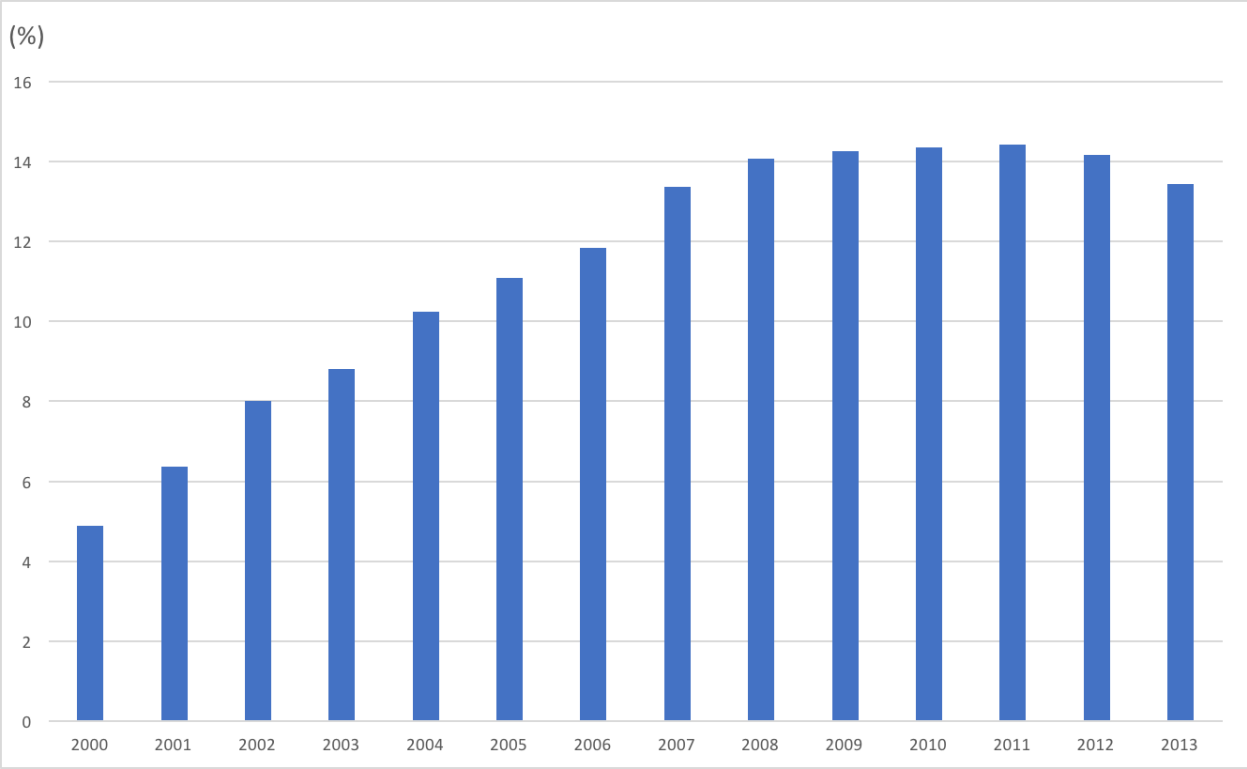
characteristics of the immigrant population. In particular, the last outcome analysed refer to the different job categories individuals are classified in and their professional situation, i.e. whether they are self-employed.

Among the European countries with a history of immigration, Spain is one of those with a more recent experience. Since the end of the 20th century, the country recorded a high level of immigration for a short period of time with an average net inflow of foreign-born individuals estimated at 500,000 people per year between 2000 and 2009 (Arango, 2013). Indeed, in Figure 1, the evolution of foreign-born population in Spain is displayed as a percentage of the total population and we noticed a remarkable rise of about 9% between 2000 and 2011. Spain is an interesting case to study, its immigration policies attracting attention internationally for how the country handled the integration of its foreign population. Unlike in some of its neighbours, the anti-immigrant sentiment among the public and political parties was restrained in Spain (Arango, 2013). Moreover, the focus on Spain to study the association between immigrants' language abilities and their situation in the labour market has become a subject of interest for researchers only more recently.

While previous studies exploited the 2007 National Immigrant Survey of Spain (NISS) to investigate the effect of language skills on earnings of immigrants [Isphording (2013), Budría & Swedberg (2015), Budría et al. (2016)], we used this dataset focusing on other labour market outcomes mentioned above. We applied an OLS estimation followed by an intrumental variable (IV) strategy to correct for potential issues of the first estimation. We found that Spanish proficiency largely enhanced the labour market success of immigrants in Spain through different outcomes. First, the effect on the probability to be employed accounts for 16.5pp applying the IV method compared to 6.1pp with OLS estimates. For low-educated individuals ($\leqslant$ 10 years of schooling), the impact is even more relevant with a parameter reaching 31.7pp. Second, the number of hours of work also increase with language skills. As for the professional characteristics, the language skills have a negative impact on the probability to be self-employed.

To this end, the second section reviews the Spanish labour market situation, the related literature and the theories about language acquisition. The data and descriptive statistics are then detailed, while the fourth section explains the identification strategy used in line with the previous literature. In section 5, the results are displayed from which we drew some conclusions in section 6.

Figure 1: Evolution of the foreign-born population (in percentage of total population) in Spain, 2000-2013



Source: International Migration Database, OECD

2 Background

For a deeper understanding, this section reviews thoroughly the characteristics of the Spanish labour market over the last decades as well as some literature review about the role of languages and the immigrant labour force. Finally, theories about language proficiency are described.

2.1 Overview of the Spanish Labour Market

While this paper investigates the impact of language skills of immigrants on their labour market performance in Spain, a brief overview of the Spanish job market is convenient for a rooted study. The recent crisis has inflicted considerable damage on several indicators of the economy in Spain.

However, the last few years have been characterised by an improvement in the Spanish economic situation and, specifically, better labour market conditions with a decreasing unemployment rate. Although Spain still presented the second highest unemployment rate among OECD countries in April 2017, it decreased by more than 8 percentage points since 2013 (Keese et al., 2017).

The pre-crisis situation in Spain was characterised by an increasing number of employed invidivuals and an unemployment rate falling off since the nineties. For instance, the unemployment rate for males dropped off slowly from 2002 to 2007 from around nine to six percentage points (Malo, 2015). Regarding the foreign-born population, from 1996 to 2005, Fernández & Ortega (2006) observed that a larger number of immigrants participated in the labour force with a participation rate 15% higher than the one of natives. The authors concluded that Spain could manage the immigration inflows mostly through allocating immigrants to temporary jobs. The latter is an issue the Spanish labour market had been struggling with as it involves a labour market characterised by job instability. The government then focused on reducing the rate of unemployment at the expense of more temporary work and thus created uncertainty within the job market. However, a reform of 2006 was implemented with the objective to thwart this temporal trend in the Spanish labour market (Rubio, 2015).

Since the beginning of the crisis, Spain had experienced critical changes in its macroeconomic and financial conditions. With a relatively long economic recession, the country's GDP was 6.3% lower in the first quarter of 2009 followed by a severe decline in employment categorizing Spain as one of the Euro member states with the highest unemployment rate (Carballo-Cruz, 2011). Indeed, between 2008 and 2012, the number of jobs decreased by 2.9 million and 66.7% of the total job decline affected young people (16-29 years old) who suffered the most from this economic downturn (Sánchez, 2012). Foreign-born population was also affected during the crisis with an immigrant unemployment reaching over 36% in the first quarter of 2012 (Arango, 2013). Those were predominantly employed in the construction sector which was highly impacted by Spain's building boom combined with the global financial crisis. Overall, the crisis widened the gap between immigrants' and natives' unemployment rates. Fellini (2017) estimated this gap for men to increase from 1.7% to 16.8% in the period 2007- 2012.

In the attempt to overcome this economic deterioration, Spain, with the help of Europe, started to take control of the situation through different reforms. For instance, a reform implemented in

2010 intended to have more work flexibility, i.e easing firing procedures, liberalising temporary work or making permanent contracts less rigid (Rubio, 2015). Under Mariano Rajoy's governance, another labour market reform took place in 2012 which reinforced the capacity of employers to respond to shocks with new conditions such as easing the justification of layoffs (Fallis, 2013). After the unemployment rate reached a peak of 26.2% in 2013, Spain's economy has been subjected to a slight improvement in labour market with 950,000 jobs created and an unemployment rate falling 5.3% between the end of 2013 and the end of 2015 (Trésor-Economics, 2016). Although pre-crisis labour market conditions would not be reached by the end of 2018, OECD projections estimate that positive recovery trend will continue in the next two years with an employment rate scoring 55.4%.

In this paper, the study reflects the pre-crisis situtation of the Spanish labour market since we used data from the National Immigrant Survey of Spain which was conducted between 2006 and 2007.

2.2 Literature Review

The language proficiency of immigrants is a subject widely discussed in the empirical literature. Initially studies tended to focus on Anglo-Saxon countries such as England, Canada or the United States. At the beginning at the nineties, Chiswick (1991) was concerned about the determinants of English language fluency among immigrants and its effects on earnings. Using data from the United States, he could conclude that reading fluency is more important than speaking fluency as a determinant of earnings. Differentiating language skills, i.e understand, speak, read and write, Carnevale et al. (2001) found that an immigrant's ability to understand spoken English is the main required skill to succeed in the labour market. Dustmann & Fabbri (2003) estimated the effect of language on earnings and employment probabilities of non-white immigrants in the UK. As expected, they found a positive influence of language proficiency on the likelihood to be employed and on earnings. Looking at a different component of the labour supply, Law (2008) studied the impact of English proficiency on the hours worked by immigrants in Australia. Results suggest a positive relationship between English fluency and hours of work.

The impact for some non-English-speaking countries has also been reviewed. Isphording et al. (2014) studied the effect of decifiency in spoken and written German on the labor market out-

comes of immigrants. For the Netherlands, Yao & Van Ours (2015) evaluated the labor market performance in terms of employment, hours of work and wages influenced by Dutch language skills of immigrants. Overall, besides Hayfron (2001) who didn't identify positive returns to Norwegian proficiency for immigrants, most of the research could reach a consensus of a positive relationship between host language skills and success in the labour market.

An increasing number of studies also centered their focus on Spain. To go further than studying the relevance of local language skills, Isphording (2013) examined the returns of foreign language abilities of immigrants in the Spanish labor market. The results indicate significant wage premia for foreign language proficiency and depend on the occupational choice. Budría & Swedberg (2014) highlight the relevance of language proficiency in the labour market exploring the determinants of this human capital component among immigrants. They could determine a positive association between multiligualism and Spanish proficiency of immigrants. Using Instrumental Variables, Budría & Swedberg (2015) also found that being proficient in Spanish increases immigrants' earnings by about 20%. Furthermore, when applying a quantile regression method, Budría et al. (2016) focused on a deeper analysis along the wage distribution and found that the returns to higher education is on average 17%. For bilingual Spanish regions, such as Catalonia, the relationship between language skills and labour market performance has been especially analysed. For instance, Rendon (2007) studied the contribution of Catalan knowledge to finding a job in Catalonia and found a significant and positive Catalan premium. Focusing on the same region of Spain, Di Paolo & Raymond (2012)'s results indicate the existence of a positive return to knowledge of Catalan, with a 18% increase in earnings for individuals who are fluent in Catalan.

Whereas most of the studies focused on the effect on earnings, we provided a distinct analysis looking at various components of the labour market performance. Using the same data as in Isphording (2013), Budría & Swedberg (2014), Budría & Swedberg (2015) and Budría et al. (2016), we considered outcomes such as employment, contract duration, hours of work, the occupational and professional situation (e.g. self-employment) to estimate their causal effects with the Spanish language proficiency of immigrants in Spain. These characteristics of the labour supply are relevant for a broader understanding of immigrants' situation in the labour market.

2.3 Language Proficiency: Theories

Language is an essential component of human capital and its importance in the labour market is crucial. From social interaction to processing information, language skills enable individuals to perform a large set of tasks and are therefore valuable in the labour market in the destination country. Immigrants should consider investing in the knowledge of destination language for better chances to reach economic success. However this investment also depends on the amount of costs this language training requires which might alter immigrants' decision in learning the local language considering all the efforts required (money and time). In this sense, Chiswick & Miller (1995) express the destination language fluency as a function of three variables: economic incentives, exposure and efficiency.

From an economic point of view, immigrants should rationally assess the benefits of acquiring the destination language. The additional income correlated with higher language skills in the destination country has been extensively investigated by economists [Chiswick (1991), Dustmann & Fabbri (2003), Budría & Swedberg (2015)]. The mechanism of this relationship arose with human capital theory suggesting that education increases the productivity of workers, through knowledge and skills, hence raising their wages (Becker, 1964). Following Chiswick (2008), language skills satisfy the three requirements for human capital, namely productive, costly to produce and embodied in the person. Another important factor for immigrants to consider when evaluating the benefits of acquiring the destination language is the probability of employment. It is reasonable to believe that the more fluent are likely to be more efficient in finding a job fitting their needs, hence reducing the length of job search (Chiswick & Hurst, 2000). Furthermore, the economic incentives of language acquisition are also related to consumption activies such as lower costs of search for consuming some products and services or getting higher quality products thanks to more information.

Chiswick & Miller (1995) differentiates three phases through which *exposure* affects immigrants' language skills: before immigration, time and intensity of exposure in the destination. Prior to immigration, the exposure depends on factors such as birthplace regions or linguistic distance. For instance, hypothetically, an immigrant from Italy might learn Spanish faster than an immigrant from China due to this lower linguistic distance between Italian and Spanish compared to Chinese. Second, the time immigrants spent in the destination country is relevant for estimat-

ing their language skills. Their exposure to the native society, hence to the language, is supposed to be longer and therefore impact the learning process. Finally, what the authors defined as "the intensity of exposure" refers to whether the immigrants' mother tongue is used in their neighborhood hence escaping from destination language. This mechanism applies also for immigrants having young children. The latter are more efficient in learning a new language hence might have a positive influence on their parents' training process.

Immigrants' language skills also depend on the efficiency of the language acquisition. For instance, better educated individuals might be more competent in learning new concepts and will therefore have a greater efficiency in language acquisition. Foreign-born population getting an education in the destination country (for immigrating at an early age) is more likely to have greater fluency. In general, immigrants with more abilities will gain higher human capital, whether it is schooling or language skills. This efficiency component might also depends on the category of immigrants. For example, those described as refugees might have less fluency in the destination language since there is less self-selection due to the nature of the migration (political persecution, etc.).

With this theoretical background, we considered the following hypotheses for the different labour market outcomes we analysed. First, we expect the language fluency of immigrants to have a positive impact on their probability of employment in Spain. In other words, with higher Spanish skills, immigrants in Spain will be more likely to find a job in the labour market. A second interesting insight we want to capture is whether the language proficiency affects immigrants' employment in terms of their contract duration. Given the importance of language proficiency in the labour market, we expect immigrants with more fluency in Spanish to acquire longer work contracts. Third, hours of work are another relevant measure of labour supply complementing employment. Reflecting the success of immigrants on the job market, we expect working hours to be positively correlated with better language skills. The occupational situation of individuals is also an outcome worth analysing . In 2012, the share of employed in non-manual highly skilled jobs in Spain accounted for 42% for natives whereas the amount was only 17% for foreign-born population, reflecting a significant gap in occupations (Fellini, 2017). We predicted that the host language proficiency is more likely to enhance the position of migrants into higher job categories. Finally, we decided to detect the effect on the professional situation of the migrants referring to

whether they are self-employed. This particular outcome has been overlooked in previous studies although it provides more details for the labour market structure of migrants. We apprehend the relation between this outcome and Spanish proficiency to be positively correlated. Indeed, individuals who have low level of language skills are less likely to be self-employed given the higher costs they must encounter related to administration or information costs. With more language skills, however, individuals would have more facilities to join self-employment.

3 Data and Descriptive Statistics

In this section, we described the sources used to build our final sample. In particular, we provided in detail the construction of the different variables of interest for the empirical model. Furthermore, we presented the summary statistics of our dataset to apprehend the sample and highlight preliminary findings about the link between language proficiency and labour market outcomes.

3.1 Data

For the empirical analysis, we used data from the National Immigrant Survey of Spain (NISS), a cross-sectional survey conducted between November 2006 and February 2007. The Spanish National Statistics Insitute carried out the data collection to provide a large scale of information about immigrants in Spain and to deepen our knowledge about the migratory phenomenon. The survey compiled socio-demographic characteristics of immigrants as well as detailed information about their situation in the labour market. In the study, the NISS defines an immigrant as an invidivual born abroad with at least 16 years old and who had lived or had the intention to stay in Spain for at least one year.

The original survey sample encompasses about 15,500 individuals but we restricted the sample for the purpose of this study. First, we only considered the population active in the labour market, i.e. individuals from 16 to 64 years old[1] who are not in education, retired or permanently disabled, as the goal of this paper is to highlight the effects of language proficiency of immigrants on their labour market outcomes. Second, we ony included men to avoid additional complications of selectivity bias with women in the labour market. Afterwards, we made more restrictions

[1] At the time of the survey, the working age population was 15-64 years old.

depending on the outcome used. First, we restriced the sample by differentiating estimations depending on whether individuals reside in bilingual regions or not. Furthermore, we only considered respondents with a level of education higher than 10 years of schooling for a first estimation followed by another one only including individuals with 10 years of schooling or less. After eliminating observations with missing values of relevant variables, we obtained a final sample of 4,492 observations.

3.1.1 Spanish Proficiency

The independent variable we want to focus on is the Spanish proficiency of immigrants. To evaluate this skill, respondents had to answer these questions (translated in English):

Thinking of what you need to communicate at work, at the bank or with the administration services, how well do you speak Spanish?

Participants could choose answers from 1-*Very well*, 2-*Well*, 3-*Sufficient*, 4-*Need to improve*. Subsequently, the Spanish proficiency variable is defined as a dummy variable equal to 1 if the immigrant is fluent in Spanish, i.e. he reported *well* or *very well*.

3.1.2 Labour market outcomes

When considering the labour market environment, it is meaningful to investigate the probability of employment for the group of interest. To construct this variable, we defined an employed individual as someone with any type of paid work. Therefore, the indicator of employment is defined as a dummy variable equals to 1 if the immigrant is employed and 0 otherwise. Respondents are also asked the duration of their working contract. This variable could be identified as a dummy variable taking the value of 1 if the working individual had a permanent contract and 0 if he had a temporary position. Analysing the number of hours immigrants work can also bring some interesting insights for the labour market performance study. In the survey, respondents were asked how many hours per week did they work. The investigation of both the level of employment and hours of work specifically studies the effects of language proficiency on the extensive and intensive margins of labour supply. This applies to whether to work and how much to work for each individual.

As mentioned previously, language proficiency can also impact the occupational choice of immigrants since for each type of occupation different language skills might be required. We first constructed a variable dividing the occupations in 4 categories: high white collar, low white collar, high blue collar, low blue collar. For clarification, the first category accounts for individuals working in positions referring to *company management and public administration* as well as *intellectual and scientific professionals and experts*. The low blue collar group represents workers with less qualifications such as *operators of facilities and machinery* as well as those without any qualifications. From this, we could create a dummy variable equals to 1 if the individual has a high white collar work position and 0 if his job refers to any other category. Additionally, we analysed the relation between the language proficiency and the job characteristic of being self-employed. This outcome could be constructed as a dummy variable taking the value of 1 if migrants work independently and 0 otherwise.

Notably, job components and working hours are only reported for employed individuals. Hence unobserved characteristics might be affecting employment and the other outcomes at the same time. For instance, migrants with more Spanish fluency may be self-selected into employment and report being in the high white collar category. Therefore, our results may not be related to language proficiency but to other unobserved characteristics. For further research, addressing the concern of non-random selection would be interesting to investigate through estimating a three-equations model for proficiency, employment and job characteristics.

3.2 Summary Statistics

Table 1 gives an overview of the characteristics of our sample split-up in three groups, including the total sample and the individuals with and without Spanish proficiency. The personal characteristics of the two group of immigrants separated by the language, present some disparaties. First, immigrants with Spanish proficiency tend to have more years of education than those without such language skills with a difference of over 1 year in our sample. Second, the proficient ones seem to arrive on average at an earlier age than the non-proficient ones which follows the reasoning of previous theories. Interestingly, the household composition also presents some differences. The number of children and adults in the household is slightly higher for non-proficient individuals suggesting that migrants with less languages skills tend to live in a larger household. The

two groups are quite similar in terms of the region of residence *(Comunidad Autonoma*) aside from the Community of Madrid in which much more Spanish proficient individuals reside. Regarding the region of origins of immigrants, Latin Americans certainly have higher Spanish proficiency than individuals coming from other regions since Spanish is the first language of most of the population in Latin America. Other proficient immigrants are mostly from Central, Northern and Southern Europe. Since the years of schooling are generally higher in these regions compared to other regions of the world, these observations highlight the intuition that higher education leads to more facility to acquire a new language. Non-proficient immigrants are more likely to come from Africa, Eastern Europe or Asia and Oceania.

Concerning the labour market performance, the two groups of immigrants demonstrate some differences. The level of employment for Spanish-proficient individuals is 9% higher compared to the non-proficient ones. Holding a permanent contract also differs between the two groups as a lower number of immigrants with non-Spanish proficiency holds a temporary work position. In terms of occupation, a larger number of white collars have fluency in Spanish and a larger number of blue collars have lower skills in the host language. Concerning the other outcomes, namely hours of work and self-employment, there is not particular difference between the two groups.

4 Identification Strategy

Assuming that measurement errors are absent and language skills are exogenous to labour market outcomes, we first represented the following analysis with OLS estimates :

$$Y_i = \beta_0 + \beta_1 L_i + \beta_2 Age_i + \beta_3 Agesq_i + \beta_4 Arrival_i + \beta_5 Cbirth_i + \beta_6 X_i + \epsilon_i \tag{1}$$

Y_i refers to an indicator of the labour market outcomes. The first indicator is the dummy variable for whether an individual is employed or unemployed. The second indicator represents the duration of the work contract defined as a dummy variable equals to 1 if the contract is permanent. For the third outcome, we analysed the logarithm of hours of work per week individuals reported in the interview, considering only positive hours, i.e. individuals who work. The fourth outcome refers to the occupation and takes the value of 1 if immigrants hold a high white collar

position. Finally, the self-employment characteristic is determined as a dummy variable taking the value of 1 if migrants are self-employed.

L_i refers to the Spanish language proficiency of respondents. Furthermore, independent variables account for some relevant background characteristics, namely Age_i and $Agesq_i$ representing the age of individuals and its square , $Arrival_i$ referring to the age at arrival of the immigrant in Spain and $Cbirth_i$ indicating the country of birth of immigrants. For those last two variables described, we included dummies for each possible age at arrival, for $Arrival_i$, and each possible country of birth, for $Cbirth_i$. Additionally, X_i is a set of control variables that we progressively expanded. In this way, we included step by step in the regression the years of schooling, the marital status, the number of kids and adults living in the same household and the region of residence. Moreover, we also estimated the same model for bilingual and non-bilingual regions, as well as for individuals with high or low education (more or less than 10 years of schooling), in order to check for complementarities . Finally, ϵ_i is the error term.

In this paper, we focused on obtaining a consistent estimate of the parameter β_1 to specifically identify the causal effects of Spanish language proficiency of immigrants on their labour market performance. However, identifiying the impact of language on socio-economic outcomes can be problematic because of the endogeneity of language skills. First, measures of language proficiency are self-reported hence causing measurement error. The latter may lead to a downward bias of the effect of languages on labour market performance. Second, language skills estimates are likely to be biased due to possible omission of variables in the regression. Indeed the previous literature have argued that both language proficiency and labour market outcomes might be correlated with unoberseved heterogeneity. For instance, ability or motivation are inner characteristics that could impact an individual's language fluency and work success. Furthermore, employed immigrants are more likely to interact with the local society and to access language courses which reverse the causal effect of language proficiency on labour market performance.

The approach adopted to address these issues followed the stream of literature which applied Instrumental Variable (IV) regressions. With this method, we could assume that there is an observable variable that affects the language proficiency but is not correlated with the empirical concerns studied above. This instrumental variable is then useful to predict exogenous levels of language skills and its actual impact. Although finding the appropriate instrument is complex due to the

constant use of languages in everyday life, previous studies offered various instruments for their analysis. For instance, Chiswick & Miller (1995) used parental education, family composition and family ties as instruments to estimate the earnings of immigrants. Their results showed a large downward bias of the OLS estimates and significant higher wages using the IV method. These findings were confirmed by Dustmann and Van Soest (2002) that applied lags of language skills and parental education as instruments to correct measurement errors. Further studies kept innovating to find more instruments for language proficiency [Bleakley & Chin (2004), Rendon (2007), Gao & Smyth (2011)].

For this study, we decided to explore two instruments for Spanish proficiency which were available in the NISS survey. Following Bleakley & Chin (2004)'s method to respond to the measurement and endogeneity problems, we used an instrument defined as the interaction between age at arrival of immigrants and whether they come from a Spanish-speaking country. Nonetheless, when applying their instrumental method, Bleakley & Chin (2004) were using data with individuals who migrated below 18 years old which is not applicable in this case since most of migrants arrived after this age (80% of our sample). Considering that a valid instrument must be correlated with the regressor, here Spanish Proficiency L_i, and uncorrelated with the error term, we considered this choice of instrument valid for our analysis. Indeed, the fact that acquiring a language at a younger age is easier than for older individuals is widely accepted. Therefore incoming population arriving at a later age tend to obtain a lower level of language proficiency than those arriving at a younger age. Apart from languages skills acquisition, younger and older incoming population differ substantially regarding, for instance, education or social inclusion. For instance, immigrating children attend the educational system of the host country and tend to have more facilities to assimilate themselves to the local population. Thereby, in line with Bleakley & Chin (2004), we included immigrants from Spanish-speaking countries to remove those effects potentially affecting labour market outcomes and not related to language assimilation. In this way, we included dummies for each potential age at migration, assuming that immigrants coming from Spanish-speaking countries experienced the same reality that immigrants from non-Spanish-speaking countries did except a new language. This choice of instrument was also adopted by Budría & Swedberg (2015) although they additionally accounted for the respondent's willingness to stay in Spain for the next five years. Isphording & Otten (2013) also applied a similar instru-

ment introducing interaction effects between the linguistic distance and the years of residence in the host country.

Applying the IV regression method involves specifying a first stage regression for Spanish proficiency:

$$L_i = \alpha_0 + \alpha_1 Arrival_i \times S_i + \beta_2 Age_i + \beta_3 Agesq_i + \beta_4 Arrival_i + \beta_5 Cbirth_i + \beta_6 X_i + \nu_i \qquad (2)$$

where $Arrival_i \times S_i$ is the instrument defined as the interaction of the age at arrival $Arrival_i$ and a variable indicating whether the immigrant comes from a Spanish-speaking country S_i. In this first stage regression, we expect the parameter of the instrument α_1 to be statistically significant and negative. Since the residuals of labour market outcomes are more likely to be correlated within region of origin, we clustered the standard errors by country of birth.

5 Results

Tables 2-6 show the OLS estimates for the effect of Spanish proficiency on the different labour market outcomes. Column 1 represents the starting point model followed by the various controls we progressively added from columns 2 to 8. According to the Table 2, having Spanish proficiency increases significantly the probability of being employed by 6.1 percentage points (pp). Once we included the number of years of education, the magnifitude of the parameter hardly decreased to 5.7pp with the same level of significance, i.e. at 1% level. Considering the household composition and the region of residence in the model, the analysis reports similar results with Spanish proficiency effects reaching 5.6pp and 5.5pp respectively. Interestingly, being married has a positive effect on being employed with a 3.2pp while having children or adults in the household do not affect this probability of employment. When we split the model for bilingual and non-bilingual regions, we found quite different estimates. Considering only non-bilingual regions, Spanish proficiency is associated with an increase in the probability of finding a job of 4.9pp, at a 5% significance level. For bilingual regions, however, the effect is higher with a parameter amounting to 7.6pp. Regarding the education that we split in high and low education, >10 and $\leqslant 10$ years of schooling respectively, the models hightlight distinct estimates. For migrants with a higher level of education, the effect of their Spanish language skills on the probability that they find a job is

0.7pp and non-significant. In this model, a curious discovery is that the number of adults in the household reduces the probability of employment by 1.4pp. For those with a lower level of education, on the contrary, the impact of Spanish proficiency on the probability to be employed reaches 11.5pp at a high significance level (1%). In other words, the effect of Spanish proficiency on the probability of employment is only relevant for lower-educated immigrants.

In Table 3, the OLS estimates are displayed to inspect the contract duration of the migrants. Holding a permanent contract is positively (4.8pp) related to Spanish proficiency but with a degree of significance at 10% level. While for the other models the impact is not meaningful, for non-bilingual regions, having Spanish fluency increases the probability of having a permanent contract by 5.1pp. As shown in Table 4, the number of hours migrants work is negatively associated with the language skills but the parameters are not significant except for the model including only non-bilingual regions of residence. Regarding the other professional characteristics of the migrants, namely the occupational category in Table 5 and the professional situation in Table 6, Spanish proficiency does not have a significant impact on those two outcomes. Table 5 only displays a 2.6pp impact on the probability of being a high-white collar for the basic model while the other models report negative and non-significant parameters. Considering the bilingual regions, table 6 reports that Spanish fluency decreases the probability of being self-employed by 6.1pp (10% significance level).

As described in the previous section, we further applied instrumental regressions to address various issues that could lead to a downward bias of our estimates. Therefore, we displayed the two-stage regression tables of this method. Table 7 reports the results of Equation 1 and highlights the outcomes of the first-stage regression, i.e. the determinants of Spanish proficiency. The instrumental variable, namely the interaction of the age at arrival $Arrival_i$ and a variable indicating whether the immigrant comes from a Spanish-speaking country, has a strongly significant effect on the language proficiency. Immigrating in Spain one year later and coming from a non-Spanish speaking country decreases the probability of acquiring Spanish proficiency by 1.2pp for the basic model. Including the controls about education, household composition and region of residence increases the parameter to 1.3pp. We noticed that, for these models, the number of children in the household does not have a significant impact on Spanish skills but the number of adults in the household has a decreasing effect on language fluency. A potential justification of this observation

is that immigrants sharing the same household would speak their native language between each other hence weaken the host language proficiency. When considering bilingual and non-bilingual regions, the impact is the same with a coefficient equals to 1.4pp at 1% significance level. Furthermore, immigrants arriving at a later age in Spain differ in their Spanish language skills depending on their education. Indeed, the negative effect of arriving at a later age on the Spanish fluency is 0.4pp higher for the lower educated individuals rather the high educated ones. While for the higher educated ones the impact is -1.1pp, for the lower educated immigrants the effect of arriving at a later age on the Spanish fluency decreases.

The IV results of the second-stage regression are represented in Tables 8 - 12 for the different labour market outcomes. Regarding employment displayed in Table 8, the impact of language fluency on the probability to be employed is higher than the effect with OLS estimates. Spanish proficiency is associated with the probability of getting a job of 16.5pp for the first model. The magnitude increases when adding the different controls suggesting that the number of schooling years, the household composition and the region of residence influence even more the impact of language skills on being employed. When spliting the sample between bilingual and non-bilingual regions of residence, the effect is even greater reaching 2.28pp. A large gap between the parameters of low and high education appears for these IV estimates. For migrants with more than 10 years of schooling, acquiring better Spanish fluency does not have a significant impact on the probability of getting a job. However, for migrants with lower years of schooling, Spanish proficiency significantly increases their chances of employment by 31.7pp. The magnitude is 20pp higher compared to OLS estimates reflecting a downward bias when using this method.

In table 9 the IV parameters for holding a permanent contract are not significantly affected by language skills. Indeed, for all models, although the magnifiture is higher, the parameters are not significant. Moreover, we identified that one additional year of schooling increases the probability of holding a permanent contract by 0.9pp in column (2) and by 0.8pp and 0.7pp for columns (3) and (4), respectively. However, the education does not have a significant effect on the contract duration for the other models. When analysing Table 10, the first two columns highlight that Spanish proficiency affects positively the number of working hours by around 11%. The impact is reversal when comparing with the OLS estimates which displayed a negative and non-significant relation. Looking at the other models, we detected a positive relationship but not

significant. Regarding the occupational job categories, Table 11 indicates a larger effect than the one of OLS estimates but with a negative and non-significant parameter for most of the model. In line with the theoretical thoughts, education plays an important role in determining the job categories. For migrants with more than 10 years of schooling, better language skills imply a negative probability to hold a high-white collar position of -37.3pp. For immigrants with lower level of education, the language fluency doesn't impact the probability of getting a job in the high-white collar category. Those results are quite ambiguous given the general belief that more educated people would hold a job position in the white collar categories. A possible explanation is that the demand for high educated people is higher than the supply. Therefore, employers would hire migrants with a high level of education to fill this gap even though the Spanish proficiency is not among their best skills. Another potential reason is that the role of the Spanish language in higher job categories is not fundamental for the purpose of the work. For instance, academic researchers, usually holding a PhD degree, are mostly required with English fluency to conduct their research. Finally, Table 12 provides a straightforward association between language proficiency and being self-employed with a negative and highly significant coefficient for all models. The first model indicates that Spanish fluency decreases the probability of being self-employed by 35.4pp. The impact decreases slightly and progressively reaching 32.4pp once we included the controls for education, household composition and region of residence. Considering only non-bilingual and bilingual regions, the effect rose to 57.4pp and 57.8pp respectively. Regarding the educational differences of migrants, the high-educated and Spanish fluent ones decrease their chances of being self-employed by 53.7pp while the impact is -24.8pp for low-educated individuals. These results are not consistent with the theoretical analysis that highlights the economic incentives of learning a new language. In this sense, migrants who want to be self-employed might face more costs associated with acquiring language skills due to administrative or searching activities.Therefore, they would be discouraged to choose this type of employment.

As mentioned in the previous section, the validity of the instrument is important for a consistent interpretation of the results. Table 8-12 reports the F-statistic to test the relevance of the instrumental variable. A F-statistic above 10 indicates that the instrument used is acceptable. For all the models and tables, the value of the F-stat reports higher values than 10 suggesting that the interaction of the age at arrival $Arrival_i$ and a variable indicating whether the immigrant comes

from a Spanish-speaking country does not appear to be a weak instrument.

6 Conclusions

The host language proficiency holds an important place in migrants' integration process in the host country. With the increasing number of incoming population in Spain, and Europe in general, the migratory phenomenon should be handled fast and efficienctly to respond to this large proportion of immigrants. Especially, the assimilation process into the labour market of the host country needs particular examination given a larger amount of unemployed migrants compared to natives.

Using the NISS of 2007, we analysed a relevant component of the integration procedure of immigrants in Spain: language fluency. We looked at the causal effect of Spanish proficiency on different labour market outcomes complementing the existing literature that focused on earnings. We first applied an OLS estimation assuming that measurement errors are absent and that language skills are exogenous to labour market outcomes. The results indicate that Spanish proficiency affects the probability of employment by 6.1pp with an increasing effect reaching 11.5pp for low-educated individuals. The OLS estimates for the other outcomes didn't display a critical relationship between the variables. Nonetheless, to address the issues related with measurement errors and endogeneity, we further developed the analysis using an intrumental (IV) estimation. That methodology allows for consistent estimates and hence a proper causal effect. Following the idea of Bleakley & Chin (2004) for the choice of instrument, we proposed an instrumental variable constructed by the interaction of the age at arrival and a variable indicating whether the immigrant comes from a Spanish-speaking country.

Regarding the level of employment, the IV estimates resulted in higher estimates than OLS's with a probability of getting a job of 16.5pp for the basic model. A relevant finding is that, Spanish proficiency significantly increases the chances of being employed by 31.7pp for low-educated migrants in opposition of high-educated ones who didn't have any relevant impact. The number of hours worked also significantly rise with language skills. Furthermore, Spanish fluency has a strong negative effect on the probability of being self-employed. The impact is even stronger for high-educated individuals.

Overall, the results suggest that Spanish proficiency enhances the labour market success of

immigrants in Spain through different outcomes.Those findings are in line with the previous literature. Furthermore, language skills are even more relevant for low-educated migrants who experienced more success in the labour market thanks to this essential human capital component. Those findings highlight the importance of focusing on individuals with lower education to address language training programs. They would benefit more in terms of employment by receiving the proper linguistic training. Interestingly, this finding is opposite when analysing the earnings, since Budría & Swedberg (2015) found that the earnings return to language skills are higher for the high-educated individuals.

This paper also presents some limitations. Concerning the use of instrument, we followed Bleakley & Chin (2004) who instrumented English skills as the interaction of arriving in the US before age 11 and a dummy for being born in a non-English speaking country. Since our data mostly encompasses migrants arriving after the age of 18, we could not properly apply their instrumental strategy which focused on childhood immigrants. Nonetheless, we believed that acquiring language skills is easier at an earlier age, even though this age range is after 18 years old. Furthermore, as mentioned in the Data section, this paper did not explore the problem of self-selection which might obstruct the interpretation of our results. When analysing job characteristics of immigrants, we only worked with employed individuals which can be self-selected. Our results might be affected by other unobserved characteristics such as ability instead of language proficiency. Therefore, estimating a three-equations model for proficiency, employment and job characteristics would be interesting for further investigation. It is also relevant to notice that the dataset used in this paper is cross-sectional hence missing the longitudinal perspective of analysing. Once a panel data survey of this type is available, it would be worthy to analyse the individuals over time and see their evolution in the labour market.

References

Arango, J. (2013). EXCEPTIONAL IN EUROPE? Spain's Experience with Immigration and Integration. *Mpi*, 22.

Becker, G. (1964). *Human Capital* (No. 2nd edition).

Bleakley, H., & Chin, A. (2004). Language Skills and Earnings : Evidence from Childhood Immigrants. *The Review of Economics and Statistics*, *86*(2), 481–496.

Budría, S., & Swedberg, P. (2014). The Impact of Multilingualism on Spanish Language Acquisition among Immigrants in Spain. (8748).

Budría, S., & Swedberg, P. (2015). THE IMPACT OF LANGUAGE PROFICIENCY ON IMMIGRANTS' EARNINGS. *Revista de Economia Aplicada*, 63–91.

Budría, S., Swedberg, P., & Fonseca, M. (2016). Returns to Schooling among Immigrants in Spain : Returns to Schooling among Immigrants in Spain : A Quantile Regression Approach. (10064).

Carballo-Cruz, F. (2011). Causes and consequences of the Spanish economic crisis: Why the recovery is taken so long? *Panoeconomicus*, *58*(3), 309–328. doi: 10.2298/PAN1103309C

Carnevale, A. P., Fry, R. A., & Lowell, B. L. (2001). Understanding, speaking, reading, writing, and earnings in the immigrant labor market. *American Economic Review*, *91*(2), 159–163.

Chiswick, B. R. (1991). Speaking, reading and earnings among low-skilled immigrants. *Journal of labor economics*, *9*(2), 149–170.

Chiswick, B. R. (2008). The Economics of Language: An Introduction and Overview. *IZA Discussion Paper, 3568*(3568), 31.

Chiswick, B. R., & Hurst, M. E. (2000). The employment, Unemployment and Unemploymen Compensation Benefits of Immigrants. *IZA Discussion Paper*, *129*.

Chiswick, B. R., & Miller, P. (1995). The endogeneity between language and earnings: International analyses. *Journal of labor economics*, *13*(2), 256–288.

Di Paolo, A., & Raymond, J. L. (2012). LANGUAGE KNOWLEDGE AND EARNINGS IN CATALONIA. *Journal of applied economics*, *15*(1), 89–118.

Dustmann, C., & Fabbri, F. (2003). Labour market performance of immigrants in the UK labour market. *Economic Journal, Royal Economic Society*, *113*, 695–717. doi: 10.1111/1468-0297.t01-1-00151

Fallis, A. (2013). the 2012 Labour Market Reform in Spain :. *OECD*, *53*(9), 1689–1699. doi: 10.1787/9789264213586-en

Fellini, I. (2017). Immigrants' labour market outcomes in Italy and Spain: Has the Southern European model disrupted during the crisis? *Migration Studies*, *6*(1). doi: 10.1093/migration/mnx029

Fernández, C., & Ortega, C. (2006). Labour Market Assimilation of Immigrants in Spain: Employment at the Expense of Bad Job-Matches? , *3*(644).

Gao, W., & Smyth, R. (2011). Economics returns to speaking standard Mandarin among migrants in China's urban labour market. *Economics of Education Review*, *30*(2), 342–352.

Hayfron, J. E. (2001). Language training , language proficiency and earnings of immigrants in Norway Language training , language proficiency and earnings of immigrants in Norway. , *1993*(October 2012), 37–41.

Isphording, I. E. (2013). Returns to Local and Foreign Language Skills. *Labour*, *27*(4), 443–461.

Isphording, I. E., & Otten, S. (2013). The Costs of Babylon - Linguistic Distance in Applied Economics. *Review of International Economics*, *21*(2), 354–369. doi: 10.1111/roie.12041

Isphording, I. E., Otten, S., & Sinning, M. (2014). The Effect of Language Deficiency on Immigrant Labor Market Outcomes in Germany. *Mimeo*.

Juchno, P., & Agafitei, M. (2017). Migrant integration 2017 edition. *eurostat*.

Keese, M., Haywood, L., & Denk, O. (2017). OECD Employment Outlook. *OECD Employment Outlook*(June), 1–2. doi: 10.1787/empl

Law, V. (2008). English Proficiency and Labour Supply of Immigrants in Australia. *Crawford School Research Paper*(12), 1–32.

Malo, M. Á. (2015). Labour Market Measures in Spain 2008 – 13 : The Crisis and Beyond Author. *International Labour office*. doi: 10.13140/RG.2.1.4540.7124

Metcalfe-Hough, V. (2015). The migration crisis? Facts, challenges and possible solutions. *Overseas Development Institute (ODI)*(Briefing Papers), 2.

Rendon, S. (2007). The Catalan premium, language and employment in Catalonia. *Journal of Population Economics*, *20*(3), 669–686.

Rubio. (2015). EL MERCADO LABORAL ESPAÑOL : RECORRIDO HISTÓRICO , COMPARATIVA EUROPEA Y ANÁLISIS DE POLÍTICA ECONÓMICA. *Facultad de ciencias sociales y empresariales (ICADE)*.

Sánchez, F. R. (2012). Youth Unemployment in Spain: Situation ans Policy Recommendations. (November).

Trésor-Economics. (2016). Spain ' s labour market reform : an initial assessment. (174).

Yao, Y., & Van Ours, J. (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics*, *34*, 76–85.

A Tables

Table 1: Summary Statistics

	Total Sample			Proficiency			Non-Proficiency		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Spanish Proficiency 4,492	0.82	0.38	3,705	1.00	0.00	787	0.00	0.00	
Personal Characteristics									
Age	4,492	36.08	9.63	3,705	36.11	9.61	787	35.93	9.77
Years of education	4,492	10.90	3.36	3,705	11.14	3.23	787	9.79	3.75
Single	4,492	0.39	0.49	3,705	0.39	0.49	787	0.38	0.49
Married	4,492	0.56	0.50	3,705	0.56	0.50	787	0.56	0.50
Widowed	4,492	0.00	0.06	3,705	0.00	0.06	787	0.01	0.07
Separated	4,492	0.02	0.14	3,705	0.02	0.14	787	0.02	0.12
Divorced	4,492	0.03	0.17	3,705	0.03	0.17	787	0.04	0.19
Number of children in household	4,492	0.48	0.96	3,705	0.45	0.94	787	0.60	1.05
Number of adults in household	4,492	2.62	1.19	3,705	2.57	1.14	787	2.82	1.38
Age at arrival	4,492	25.35	11.38	3,705	24.16	11.38	787	30.92	9.63
Autonomous community									
Andalusia	4,492	0.07	0.25	3,705	0.07	0.25	787	0.08	0.27
Aragon	4,492	0.04	0.20	3,705	0.04	0.19	787	0.06	0.24
Asturias	4,492	0.04	0.18	3,705	0.04	0.19	787	0.02	0.14
Baleares Islands	4,492	0.07	0.25	3,705	0.07	0.25	787	0.07	0.25
Canarias Islands	4,492	0.05	0.21	3,705	0.05	0.21	787	0.06	0.23
Cantabria	4,492	0.03	0.16	3,705	0.03	0.17	787	0.02	0.13
Castile Leon	4,492	0.04	0.20	3,705	0.04	0.20	787	0.04	0.19
Castile-La Mancha	4,492	0.05	0.21	3,705	0.05	0.21	787	0.05	0.23
Catalonia	4,492	0.12	0.33	3,705	0.12	0.33	787	0.10	0.30
Community of Valencia	4,492	0.09	0.29	3,705	0.08	0.28	787	0.11	0.32
Extremadura	4,492	0.02	0.14	3,705	0.02	0.14	787	0.03	0.17
Galicia	4,492	0.04	0.19	3,705	0.04	0.20	787	0.03	0.18
Community of Madrid	4,492	0.13	0.34	3,705	0.14	0.35	787	0.08	0.28

Table 1: Summary Statistics (continuation)

	Total Sample			Proficiency			Non-Proficiency		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Region of Murcia	4,492	0.07	0.25	3,705	0.06	0.24	787	0.09	0.28
Navarre	4,492	0.08	0.27	3,705	0.09	0.28	787	0.05	0.23
Basque country	4,492	0.04	0.19	3,705	0.04	0.19	787	0.03	0.17
La Rioja	4,492	0.03	0.18	3,705	0.02	0.16	787	0.06	0.25
Ceuta	4,492	0.00	0.06	3,705	0.00	0.06	787	0.00	0.05
Melilla	4,492	0.00	0.07	3,705	0.00	0.07	787	0.01	0.07
Region of origin									
Latin America	4,476	0.38	0.49	3,690	0.46	0.50	786	0.03	0.16
Maghreb	4,476	0.14	0.35	3,690	0.12	0.33	786	0.23	0.42
Rest of Africa	4,476	0.04	0.20	3,690	0.03	0.17	786	0.09	0.29
Eastern Europe	4,476	0.17	0.38	3,690	0.13	0.34	786	0.36	0.48
Western, northern and Southern Europe	4,476	0.22	0.41	3,690	0.22	0.41	786	0.20	0.40
Northern America	4,476	0.01	0.09	3,690	0.01	0.10	786	0.00	0.06
Asia and Oceania	4,476	0.04	0.19	3,690	0.03	0.16	786	0.10	0.29
Labour Market Outcomes									
Level of employment	4,492	0.87	0.34	3,705	0.88	0.32	787	0.80	0.40
Contract Duration	3,208	0.55	0.50	2,701	0.58	0.49	507	0.40	0.49
Hours of work	3,803	3.75	0.26	3,191	3.75	0.26	612	3.75	0.26
Occupation									
High white collar	3,883	0.15	0.36	3,253	0.17	0.37	630	0.09	0.28
Low white collar	3,883	0.10	0.30	3,253	0.11	0.31	630	0.05	0.22
High blue collar	3,883	0.42	0.49	3,253	0.42	0.49	630	0.45	0.50
Low blue collar	3,883	0.32	0.47	3,253	0.31	0.46	630	0.41	0.49
Self-employed	3,847	0.14	0.35	3,231	0.15	0.35	616	0.14	0.34

Table 2: Employment, OLS estimates

		Controls						
	(1) First Model	(2) Schooling years	(3) Household composition	(4) Region of Residence	(5) Non-bilingual regions	(6) Bilingual regions	(7) High education	(8) Low education
Spanish proficiency	0.061***	0.057***	0.056***	0.055***	0.049**	0.076**	0.007	0.115***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.022)	(0.034)	(0.023)	(0.028)
Age	0.032***	0.031***	0.027***	0.027***	0.018***	0.040***	0.019***	0.036***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.009)	(0.007)	(0.009)
Age2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000**	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.003*	0.003	0.003	0.005**	-0.001	0.005	0.002
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)
Married			0.032***	0.031***	0.046***	0.008	0.037***	0.024
			(0.011)	(0.011)	(0.014)	(0.021)	(0.014)	(0.021)
Number of children in household			-0.001	-0.000	0.003	-0.001	0.004	-0.008
			(0.007)	(0.007)	(0.008)	(0.012)	(0.009)	(0.011)
Number of adults in household			-0.007	-0.008	-0.009	-0.002	-0.014**	-0.001
			(0.005)	(0.005)	(0.006)	(0.010)	(0.006)	(0.008)
Constant	0.449***	0.422***	0.496***	0.481***	0.635***	0.238	0.669***	0.325*
	(0.106)	(0.107)	(0.114)	(0.120)	(0.150)	(0.206)	(0.154)	(0.181)
N	4492	4492	4492	4492	2903	1589	2676	1816
adj. R^2	0.039	0.040	0.041	0.046	0.046	0.053	0.041	0.058

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 3: Contract Duration, OLS estimates

		Controls						
	(1) First Model	(2) Schooling years	(3) Household composition	(4) Region of Residence	(5) Non-bilingual regions	(6) Bilingual regions	(7) High education	(8) Low education
Spanish Proficiency	0.048*	0.038	0.035	0.029	0.051*	-0.012	0.017	0.047
	(0.027)	(0.027)	(0.027)	(0.027)	(0.030)	(0.053)	(0.038)	(0.038)
Age	0.032***	0.031***	0.024***	0.027***	0.021**	0.031**	0.026**	0.023*
	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)	(0.015)	(0.010)	(0.013)
Age2	-0.000**	-0.000**	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.010***	0.009***	0.008***	0.010***	0.008	0.014*	0.007
		(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.008)	(0.006)
Married			0.046**	0.046**	0.058***	0.026	0.060**	0.030
			(0.020)	(0.020)	(0.022)	(0.035)	(0.025)	(0.031)
Number of children in household			-0.011	-0.009	-0.011	-0.010	-0.017	-0.016
			(0.011)	(0.011)	(0.013)	(0.022)	(0.016)	(0.016)
Number of adults in household			-0.018**	-0.016**	-0.014*	-0.034**	-0.017*	-0.013
			(0.008)	(0.008)	(0.008)	(0.015)	(0.010)	(0.011)
Constant	0.446**	0.345*	0.514***	0.322*	0.590**	0.709**	0.126	0.925***
	(0.174)	(0.177)	(0.187)	(0.185)	(0.246)	(0.308)	(0.299)	(0.319)
N	3208	3208	3208	3208	2564	1063	1915	1513
adj. R^2	0.151	0.155	0.157	0.166	0.171	0.127	0.161	0.138

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 4: Hours of work, OLS estimates

		Controls						
	(1) First Model	(2) Schooling years	(3) Household composition	(4) Region of Residence	(5) Non-bilingual regions	(6) Bilingual regions	(7) High education	(8) Low education
Spanish proficiency	-0.008	-0.007	-0.008	-0.005	-0.032**	-0.000	-0.029	-0.002
	(0.013)	(0.013)	(0.013)	(0.013)	(0.015)	(0.027)	(0.019)	(0.017)
Age	0.010**	0.010**	0.008*	0.007	0.018***	0.012	0.002	0.019**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.008)	(0.006)	(0.008)
Age2	-0.000**	-0.000**	-0.000*	-0.000	-0.000***	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		-0.001	-0.001	-0.001	-0.001	-0.002	-0.004	-0.001
		(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.003)
Married			0.016	0.017*	0.023**	0.003	0.026**	0.004
			(0.010)	(0.010)	(0.011)	(0.018)	(0.013)	(0.015)
Number of children in household			0.010*	0.009*	0.007	0.018	-0.001	0.018**
			(0.005)	(0.005)	(0.005)	(0.013)	(0.007)	(0.007)
Number of adults in household			-0.002	-0.003	0.000	-0.005	-0.010**	0.004
			(0.004)	(0.004)	(0.004)	(0.007)	(0.005)	(0.006)
Constant	3.463***	3.471***	3.506***	3.561***	3.757***	3.402***	3.704***	3.671***
	(0.086)	(0.088)	(0.092)	(0.102)	(0.140)	(0.169)	(0.155)	(0.190)
N	3803	3803	3803	3803	2970	1311	2295	1760
adj. R^2	0.031	0.030	0.032	0.035	0.048	0.034	0.030	0.059

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 5: Occupation, OLS estimates

		Controls						
	(1) First Model	(2) Schooling years	(3) Household composition	(4) Region of Residence	(5) Non-bilingual regions	(6) Bilingual regions	(7) High education	(8) Low education
Spanish Proficiency	0.026*	-0.002	-0.004	-0.004	-0.007	-0.036	-0.002	-0.008
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.028)	(0.021)	(0.014)
Age	-0.006	-0.007	-0.010*	-0.010*	-0.010*	-0.008	-0.008	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.010)	(0.008)	(0.006)
Age2	0.000***	0.000***	0.000***	0.000***	0.000***	0.000*	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.028***	0.027***	0.027***	0.027***	0.027***	0.088***	0.004**
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.002)
Married			0.005	0.005	0.002	0.005	0.012	0.013
			(0.012)	(0.012)	(0.013)	(0.020)	(0.016)	(0.012)
Number of children in household			-0.001	-0.002	-0.007	0.005	0.002	-0.010**
			(0.005)	(0.005)	(0.006)	(0.011)	(0.009)	(0.004)
Number of adults in household			-0.012***	-0.013***	-0.013***	-0.011	-0.015***	-0.005
			(0.004)	(0.004)	(0.004)	(0.008)	(0.006)	(0.004)
Constant	0.507	0.232	0.315	0.326	0.785***	0.168	-0.347	0.828***
	(0.362)	(0.372)	(0.371)	(0.373)	(0.180)	(0.493)	(0.426)	(0.107)
N	3883	3883	3883	3883	3015	1349	2351	1785
adj. R^2	0.239	0.292	0.292	0.292	0.304	0.275	0.368	0.096

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 6: Professional Situation, OLS estimates

		Controls						
	(1) First Model	(2) Schooling years	(3) Household composition	(4) Region of Residence	(5) Non-bilingual regions	(6) Bilingual regions	(7) High education	(8) Low education
Spanish Proficiency	0.001	-0.003	-0.003	-0.004	0.009	-0.061*	-0.020	0.013
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.034)	(0.025)	(0.022)
Age	0.000	0.000	0.002	0.003	0.004	0.002	0.007	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.011)	(0.008)	(0.009)
Age2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.004**	0.004**	0.004**	0.002	0.007**	0.007	0.002
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.003)
Married			-0.030**	-0.030**	-0.008	-0.063***	-0.037**	-0.007
			(0.013)	(0.013)	(0.014)	(0.023)	(0.018)	(0.018)
Number of children in household			-0.001	-0.001	-0.003	0.007	0.000	-0.007
			(0.006)	(0.007)	(0.006)	(0.013)	(0.010)	(0.008)
Number of adults in household			-0.008*	-0.007	-0.007	-0.005	-0.011	-0.002
			(0.004)	(0.005)	(0.004)	(0.010)	(0.007)	(0.006)
Constant	0.385	0.345	0.355	0.322	-0.180	0.385	0.296	0.217
	(0.354)	(0.356)	(0.351)	(0.356)	(0.182)	(0.519)	(0.432)	(0.302)
N	3847	3847	3847	3847	2993	1325	2335	1761
adj. R^2	0.116	0.117	0.118	0.120	0.122	0.127	0.112	0.122

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 7: Determinants of Spanish Proficiency, First-stage regression

		Controls						
	(1) First Model	(2) Schooling years	(3) Household composition	(4) Region of Residence	(5) Non-bilingual regions	(6) Bilingual regions	(7) High education	(8) Low education
Instrument ($Arrival_i \times S_i$)	-0.012***	-0.013***	-0.013***	-0.013***	-0.014***	-0.014***	-0.011***	-0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.014**	0.012**	0.009	0.010	0.003	0.003	0.013	0.012*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.006)
Age2	-0.000	-0.000	-0.000	-0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.014***	0.013***	0.013***	0.010***	0.010***	0.012**	0.016***
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Married			0.010	0.009	0.014	0.014	0.026*	-0.022
			(0.015)	(0.015)	(0.020)	(0.020)	(0.015)	(0.022)
Number of children in household			-0.003	-0.003	0.004	0.004	0.000	-0.004
			(0.007)	(0.007)	(0.011)	(0.011)	(0.007)	(0.010)
Number of adults in household			-0.016***	-0.016***	-0.024**	-0.024**	-0.010*	-0.018*
			(0.006)	(0.006)	(0.009)	(0.009)	(0.006)	(0.010)
Constant	0.382***	0.251*	0.353***	0.323**	0.427**	0.427**	0.096	0.821***
	(0.139)	(0.148)	(0.134)	(0.141)	(0.181)	(0.181)	(0.220)	(0.102)
N	4492	4492	4492	4492	1589	1589	2676	1816
adj. R^2	0.297	0.309	0.311	0.315	0.311	0.344	0.290	0.359

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies; c) Instrument defined as the interaction of the age at arrival $Arrival_i$ and coming from a Spanish-speaking country

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

Table 8: Employment, IV estimates

		Controls						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First Model	Schooling years	Household composition	Region of Residence	Non-bilingual regions	Bilingual regions	High education	Low education
Spanish Proficiency	0.165**	0.168**	0.171**	0.181**	0.228**	0.228**	0.061	0.317***
	(0.075)	(0.074)	(0.072)	(0.070)	(0.098)	(0.098)	(0.120)	(0.105)
Age	0.030***	0.030***	0.026***	0.026***	0.039***	0.039***	0.018*	0.033***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.008)
Age2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000*	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.001	0.001	0.001	-0.003	-0.003	0.004	-0.001
		(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.005)	(0.005)
Married			0.031**	0.030**	0.006	0.006	0.036**	0.030
			(0.012)	(0.013)	(0.021)	(0.021)	(0.016)	(0.021)
Number of children in household			-0.001	-0.001	-0.003	-0.003	0.004	-0.009
			(0.007)	(0.008)	(0.013)	(0.013)	(0.008)	(0.011)
Number of adults in household			-0.005	-0.006*	0.001	0.001	-0.013**	0.002
			(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.007)
Constant	0.419***	0.403***	0.466***	0.452***	0.198	0.198	0.669***	-0.710***
	(0.126)	(0.132)	(0.140)	(0.159)	(0.226)	(0.226)	(0.228)	(0.195)
N	4492	4492	4492	4492	1589	1589	2676	1816
adj. R^2	0.029	0.028	0.029	0.032	0.035	0.036	0.039	0.022
F	26372.452	1.08e+06	1.68e+07	6.22e+05	3.97e+05	9.98e+05	2.89e+07	573.487

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 9: Contract Duration, IV estimates

		Controls						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First Model	Schooling years	Household composition	Region of Residence	Non-bilingual regions	Bilingual regions	High education	Low education
Spanish Proficiency	0.074	0.098	0.135	0.125	0.313	0.295	0.115	0.262
	(0.142)	(0.134)	(0.121)	(0.122)	(0.192)	(0.206)	(0.146)	(0.169)
Age	0.032***	0.031***	0.022***	0.025***	0.035***	0.036***	0.025**	0.021***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.012)	(0.010)	(0.010)	(0.007)
Age2	-0.000**	-0.000**	-0.000	-0.000*	-0.000*	-0.000**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.009***	0.008**	0.007**	0.006	0.005	0.012	0.004
		(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.008)	(0.005)
Married			0.045***	0.046***	0.028	0.025	0.060**	0.033
			(0.016)	(0.014)	(0.033)	(0.036)	(0.025)	(0.025)
Number of children in household			-0.012	-0.010	-0.015	-0.008	-0.013	-0.017
			(0.012)	(0.011)	(0.015)	(0.018)	(0.023)	(0.014)
Number of adults in household			-0.017	-0.015	-0.026**	-0.023**	-0.014	-0.011
			(0.012)	(0.012)	(0.010)	(0.011)	(0.014)	(0.014)
Constant	0.449***	0.359**	0.528***	0.344**	0.322	0.394*	0.072	0.431**
	(0.153)	(0.144)	(0.187)	(0.171)	(0.217)	(0.204)	(0.263)	(0.204)
N	3208	3208	3208	3208	1263	1063	1915	1513
adj. R^2	0.151	0.153	0.153	0.162	0.169	0.103	0.170	0.118
F	1249.682	3895.428	4177.001	44535.942	15443.071	20401.285	3.41e+05	6537.235

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 10: Hours of work, IV estimates

		Controls						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First Model	Schooling years	Household composition	Region of Residence	Non-bilingual regions	Bilingual regions	High education	Low education
Spanish Proficiency	0.114*	0.109*	0.096	0.093	0.061	0.191	0.126	0.060
	(0.069)	(0.066)	(0.064)	(0.061)	(0.134)	(0.118)	(0.088)	(0.092)
Age	0.008**	0.009***	0.007**	0.006**	0.017***	0.011*	-0.000	0.018***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.005)	(0.006)
Age2	-0.000**	-0.000***	-0.000**	-0.000*	-0.000**	-0.000	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		-0.002	-0.002	-0.002	-0.005*	-0.003	-0.005	-0.002
		(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)	(0.003)
Married			0.014	0.015*	0.014	-0.002	0.023*	0.004
			(0.009)	(0.009)	(0.014)	(0.015)	(0.012)	(0.011)
Number of children in household			0.009**	0.008**	0.008	0.017	-0.002	0.017**
			(0.004)	(0.004)	(0.011)	(0.010)	(0.007)	(0.007)
Number of adults in household			-0.001	-0.002	-0.001	-0.001	-0.008**	0.005
			(0.003)	(0.003)	(0.006)	(0.007)	(0.004)	(0.004)
Constant	3.426***	3.448***	3.476***	3.539***	3.330***	3.324***	3.702***	3.664***
	(0.085)	(0.083)	(0.089)	(0.086)	(0.148)	(0.150)	(0.137)	(0.162)
N	3803	3803	3803	3803	1543	1311	2295	1760
adj. R^2	0.009	0.011	0.016	0.021	0.032	-0.009	-0.000	0.053
F	4861.409	1849.596	12514.411	1.37e+07	82922.031	20289.738	3.71e+06	13724.674

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 11: Occupation, IV estimates

		Controls						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First Model	Schooling years	Household composition	Region of Residence	Non-bilingual regions	Bilingual regions	High education	Low education
Spanish Proficiency	-0.310	-0.234	-0.224	-0.233	-0.222	-0.226	-0.373*	-0.056
	(0.198)	(0.162)	(0.160)	(0.164)	(0.160)	(0.154)	(0.219)	(0.102)
Age	-0.000	-0.004	-0.007	-0.007	-0.011*	-0.006	-0.002	-0.000
	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.007)	(0.008)	(0.004)
Age2	0.000***	0.000***	0.000***	0.000***	0.000***	0.000**	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.030***	0.030***	0.029***	0.028***	0.028***	0.092***	0.005***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.011)	(0.002)
Married			0.008	0.008	0.027*	0.010	0.020	0.013
			(0.014)	(0.014)	(0.015)	(0.018)	(0.017)	(0.011)
Number of children in household			0.001	0.000	0.008	0.006	0.006	-0.010**
			(0.004)	(0.004)	(0.009)	(0.010)	(0.008)	(0.005)
Number of adults in household			-0.014***	-0.016***	-0.014**	-0.016*	-0.018***	-0.006
			(0.004)	(0.004)	(0.007)	(0.009)	(0.006)	(0.004)
Constant	0.600***	0.271*	0.372**	0.376**	0.272	0.223	-0.352	-0.075
	(0.200)	(0.156)	(0.166)	(0.165)	(0.355)	(0.376)	(0.268)	(0.102)
N	3883	3883	3883	3883	1589	1349	2351	1816
adj. R^2	0.152	0.251	0.256	0.253	0.311	0.251	0.298	0.022
F	50419.532	26758.002	9921.918	1.77e+05	3.59e+06	5.62e+05	2.11e+06	573.487

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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

Table 12: Professional Situation, IV estimates

		Controls						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First Model	Schooling years	Household composition	Region of Residence	Non-bilingual regions	Bilingual regions	High education	Low education
Spanish Proficiency	-0.354***	-0.335***	-0.326***	-0.324***	-0.574***	-0.578***	-0.537***	-0.248**
	(0.126)	(0.121)	(0.119)	(0.121)	(0.214)	(0.202)	(0.183)	(0.114)
Age	0.006	0.005	0.006	0.007	0.003	0.005	0.015	-0.004
	(0.006)	(0.006)	(0.006)	(0.005)	(0.010)	(0.012)	(0.010)	(0.008)
Age2	0.000	0.000	0.000	0.000	0.000	0.000	-0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling		0.008***	0.007***	0.007***	0.007**	0.010***	0.012**	0.005*
		(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)	(0.003)
Married			-0.026**	-0.026**	-0.026	-0.050**	-0.027*	-0.010
			(0.013)	(0.012)	(0.020)	(0.020)	(0.016)	(0.020)
Number of children in household			0.001	0.001	0.002	0.011	0.005	-0.003
			(0.006)	(0.006)	(0.010)	(0.011)	(0.010)	(0.006)
Number of adults in household			-0.011***	-0.010***	-0.015*	-0.017*	-0.014**	-0.004
			(0.004)	(0.003)	(0.009)	(0.010)	(0.006)	(0.006)
Constant	0.508**	0.424**	0.460**	0.412**	0.660*	0.557	0.362	0.052
	(0.204)	(0.197)	(0.196)	(0.203)	(0.378)	(0.407)	(0.332)	(0.242)
N	3847	3847	3847	3847	1558	1325	2335	1761
adj. R^2	0.015	0.030	0.036	0.040	0.101	-0.046	-0.058	0.047
F	614.130	972.870	7369.677	1.86e+05	1.66e+07	3.31e+05	1.53e+06	15469.872

a) Standard errors clustered at the country of birth level in parentheses; b) All models control for age at arrival dummies and country of birth dummies

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