

Master Degree in Economic Development and Growth
2021 - 2022

Master Thesis

Co-ethnic density and economic outcomes:
A microeconometric analysis of how residential
settlement patterns impact immigrant wages
using the Spanish National Survey of
Immigrants

Madeline Swarr

Prof. Raquel Carrasco Perea

AVOID PLAGIARISM

The University uses the **Turnitin Feedback Studio** program within the Aula Global for the delivery of student work. This program compares the originality of the work delivered by each student with millions of electronic resources and detects those parts of the text that are copied and pasted. Plagiarizing in a TFM is considered a **Serious Misconduct**, and may result in permanent expulsion from the University.



This work is licensed under Creative Commons **Attribution – Non Commercial – Non Derivatives**

Abstract

Social science has always been concerned with understanding the meso-level forces that connect individuals to the whole. In the case of the foreign immigrant, this integration into a new society, often characterized by differences in institutions, language, culture, and socioeconomic structure, is a long and arduous process that those of the native population do not always face, and if they do, do not face as intensely.

This paper defends the idea that pre-established migratory networks affect the residential settlement patterns of immigrants in regions where others from their home country tend to settle. While these networks are intuitively helpful in providing resources and protecting against the psychological shocks of such displacement, they can also help explain why those living in provinces with many other co-ethnics experience differential economic outcomes than those counterparts living in less concentrated regions. This analysis uses the 2007 National Survey of Immigrants and municipal registry data from Spain to empirically assess the relationship between residential co-ethnic density and wage differentials of immigrants who arrived to Spain between 1998 and 2007. I find evidence that this relationship varies depending on characteristics of the migratory network.

Contents

| | | |
|----------|---|-----------|
| 1 | INTRODUCTION | 1 |
| 1.1 | Theoretical Framework | 2 |
| 1.2 | Contextual Motivation | 4 |
| 2 | DATA | 5 |
| 2.1 | <i>Encuesta nacional de inmigrantes</i> (ENI) from 2007 | 5 |
| 2.1.1 | Response Variable: Log of hourly wage | 9 |
| 2.2 | <i>Estadística del Padrón continuo</i> (<i>Padrón</i> data) from 1998 - 2007 | 10 |
| 2.2.1 | Explanatory Variable: Co-ethnic density | 11 |
| 3 | METHODOLOGY | 12 |
| 4 | RESULTS | 15 |
| 4.1 | Estimating Effect of Co-Ethnic Density via OLS | 15 |
| 4.2 | Estimating Effect of Co-Ethnic Density (Absolute and Relative) via 2SLS . . | 16 |
| 4.2.1 | Endogeneity | 20 |
| 4.2.2 | IV Regression | 23 |
| 5 | CONCLUSION | 26 |
| 6 | Appendix | |

7 References

1 INTRODUCTION

Simon Kuznets observed in his early studies on the economic lives of migrants that “all minorities are characterized, at any given time, by an occupational structure distinctly narrower than that of the total population and the majority” (Kuznets, 1960, p. 14). The concept of “ethnic penalty” in immigration studies describes the fact that a meritocratic society is far from reality for these populations, and that the degree of penalty on wages and labor mobility depends on the ethnic group being referenced (Hasmath, 2012). Therefore, the classic determinants of economic well-being for immigrants do not have the same effects as for those of the native population, and these effects may differ between immigrant groups.

Many empirical studies have shown that immigrants experience lower returns to human capital than natives do, especially for education or experience attained in origin countries that are much different than host countries in terms of economic development, labor market structure, language, and institutions (see Friedberg (2000), Kanas & Van Tubergen (2009), and Le-Quang & Vallizadeh (2022) for examples). Therefore, the occupational structure is most narrow for immigrants in the years directly following arrival in the destination country when little opportunity has been given to accumulate and assimilate the human capital necessary to participate fully in the local labor market. Given the restrictions placed on newly-arrived immigrants, the impact that migrant networks can have on integration, especially in the first few years, becomes that much greater. Even if social ties are not leveraged explicitly for personal gain, the indirect influence that networks can have on protection against discrimination, information sharing, and trust can indirectly affect one’s rate of skill acquisition, job prospects, salary, and emotional well-being, as well as play a critical role in the initial decision of whether to migrate or not (Chiswick, Lee, & Miller, 2001) (Kerr & Mandorff, 2015). While migratory networks undoubtedly provide access to resources that ease the transition into a new country, it is less clear of whether the geographical concentration of immigrant groups by ethnicity that these networks bring about has an impact on the economic lives of immigrants, and if so, whether this impact is positive or negative and to what extent.

While much of the debate surrounding economic assimilation has centered around measuring the rate of convergence between native and immigrant wages over time, or economic *assimilation*, more recent research has focused on understanding what leads to differences in economic *well-being* between cohorts of immigrants. This paper not only investigates the effect of co-ethnic residential choices on hourly wages of recent immigrants to Spain, but it also sheds light on how migrant networks, actors in the settlement decision process, may be determining this effect. Utilizing detailed micro data from the 2007 National Survey of Immigrants (*Encuesta nacional de inmigrantes*) and linking them to population data from the municipal registry database (*Estadística del Padrón continuo*) to draw correlations between immigrant wages and residential co-ethnic density, I find the effect of living in a province

with a higher percentage of co-ethnics to be consistently negative and robust after controlling for certain individual-level characteristics and other fixed effects such as location and country of origin. However, I also find evidence that relative concentration of co-ethnics has the opposite effect. These conclusions remain unchanged even after addressing endogeneity issues that arise when studying voluntary migration.

The rest of the paper is structured as follows: the remainder of Section 1 discusses the theoretical framework of the study and contextualizes the debate within Spain. Section 2 describes the data and study parameters used for the empirical analysis. Section 3 outlines the wage estimation methodology. Section 4 presents the results for both the primary and secondary analyses. Finally, the paper concludes with a general remarks in Section 5.

1.1 Theoretical Framework

Many of the neoclassical models of migration predominant in the literature over the last century stem from the framework laid out by British geographer Ernst Ravenstein in his 1885 address to the Royal Statistical Society of London, *The Laws of Migration*. What may be better referred to as a set of eleven hypotheses as to why people do or do not migrate has been synthesized over time into a set of “push” and “pull” factors, the former representing those conditions in one’s present location that make life difficult or unappealing, and the latter representing those conditions in a potential new location that are attractive or more desirable in comparison (Dorigo & Tobler, 1983). Accepting the assumption that humans are utility maximizers, the migratory decision is thus understood as an economic one, a simple cost-benefit analysis to see if the pros of emigrating versus staying outweigh the cons. In support of this idea, much economic theory and empirical studies have shown that wage and employment condition differentials are key determinants in predicting the unforced flow of migrants between specific origin-destination dyads (Harris & Todaro, 1970) (Rabe & Taylor, 2012). However, if these were the only determinants, we would expect to see a much greater proportion of the world’s population living abroad than the mere 3% we see today given the current economic divide between developed and developing countries (McAuliffe & Khadria, 2020).

Why then do so many people choose to stay home rather than making the big leap? Not only can the physical and psychological costs of uprooting and replanting oneself in a foreign country be immense, but political, cultural, and informational barriers may make it impossible for one to even envision what may be waiting for them on the other side. While advances in transportation technology and changes to immigration policy also have an effect on the direction and magnitude of migratory flows, increasing emphasis in modern migration discourse has been placed on the role of co-ethnic migratory networks on the migratory process. Chain migration, which MacDonald & MacDonald (1964) define as the process by which

“prospective migrants learn of opportunities, are provided with transportation, and have initial accommodation and employment arranged *by means of primary social relationships with previous migrants*” help explain why settlement processes of immigrants, once arrived in their destination, result in relative concentration by ethnicity or foreign nationality in specific regions of the host country (p. 84). These ethnic enclaves, as they are often referred to, have been the subject of great debate amongst scholars and policymakers alike about whether they are beneficial or harmful to the economic integration of the immigrant.

On one hand, the enclave can represent a “warm embrace” in what may otherwise be a rather hostile environment to a recent transplant (Borjas, 2000). The importance of social capital amongst immigrants for correcting for information imbalances, resource scarcity, cultural displacement, restricted rights, and discrimination has been documented extensively in the literature (see Massey & España (1987), Munshi (2003), and Epstein & Gang (2010) for examples). Geographic clustering of immigrant groups by ethnicity can provide valuable economic support through job contacts, as well as create jobs directly if groups decide to establish local economies that serve their own “ethnic market” (Portes, 1981, p. 291). Other studies have shown that co-ethnic networks can breed productivity, entrepreneurship, skills acquisition, and innovation through trust and information-sharing that arises from a sense of shared identity and similar tastes (see Light, Kwun, & Zhong (1990), Ozgen, Nijkamp, & Poot (2012), and Kerr & Mandorff (2015) for examples). On the other hand, it is also argued that ethnic clustering results in, or is a result of, segregation from and discrimination by the institutional systems of the host society at large that push these groups into the secondary sector characterized by temporary contracts, minimum wage, poor working conditions, and little opportunity for promotion (see Doeringer & Piore (1975) for more). Ethnic enclaves can thus act as a double-edged sword – while providing an easy gateway into a new country, they can also hinder job mobility by restricting access to the more lucrative jobs of the primary labor market, reduce incentives to build competitive skills like local language ability, and create greater job competition and depress wages by the extent to which labor is substitutable within and free-flowing into the enclave (Borjas, 1994).

Since the mechanisms are theoretically various and directionally ambiguous, researches have turned to empirics to estimate the overall net effect of co-ethnic density on economic outcomes with various findings that differ by context and situation. Portes & Bach (1985) have studied extensively the situation of Cuban immigrants in Miami, Florida and have generally found greater returns for Cuban-owned businesses and their co-ethnic employees than for their counterparts working outside of the ethnic enclave, while Sanders & Nee (1987) have argued that it is only ethnic business owners who benefit from the solidarity and insularity of the enclave. Borjas (2000) tracks immigrant cohorts in 1980 and 1990 US Census data and find that groups who settled in metropolitan regions with higher concentration of co-ethnics experienced lower wage growth over time in comparison with their counterparts who settled

in less concentrated regions.

This paper contributes to the discussion above by attempting to isolate the independent effect of residential co-ethnic concentration on wages for recent immigrants to Spain. While initial results suggest that this effect is negative and significant, closer analysis reveals that discernment between certain qualities of the co-ethnic network may be essential to further decompose this effect.

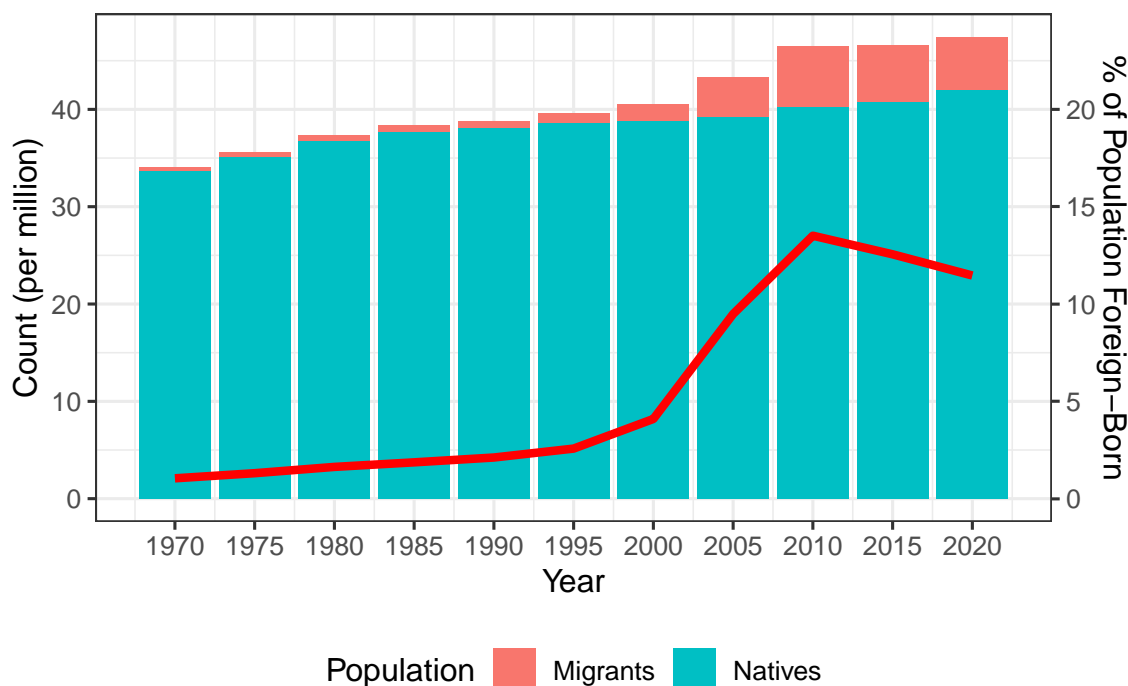
1.2 Contextual Motivation

While many studies on migrant concentration and economic well-being have been situated in countries with long histories as migrant destinations such as the United States, Canada, and Australia, research has increasingly focused on the Spanish context given the fact that it is now home to one of the largest immigrant populations in the world.¹ According to the IOM's World Migration Report, Spain was Europe's sixth most popular migrant destination in 2019 with nearly 13% of its population being made up of foreign-born individuals (McAuliffe & Khadria, 2020). Morocco, Romania, United Kingdom, and Colombia represent the top countries of origin as of the second quarter of 2021 (INE, 2021). However, all of these estimates vary due to the counting issues that arise from unregistered immigrants, as well as how an immigrant is defined. For example, an immigrant may be defined as anyone who is foreign-born (who may or may not have Spanish nationality), or anyone who possesses foreign citizenship. The estimates also vary widely depending on the geographical region within Spain. Nearly five hundred municipalities in 2022 (6%) reported populations with over 20% of all residents born in a different country (INE, 2022).

Spain as net importer of migrants is a trend that picked up speed around the turn of the twentieth century as a response to the prolonged period of growth the Spanish economy was experiencing at the time. Unemployment rates decreased drastically from nearly 25% in 1994 to record-low of around 8% at the beginning of 2007. Increased foreign investments in real estate paired with an increasingly educated but thinning youth labor force pried open a huge labor demand in the construction, agriculture, and hospitality sectors that many immigrants were able to fill. This period was also characterized by increases in female labor force participation rates, a trend further fueled by the large inflow of immigrants, mainly female and originating from Andean countries, that were meeting the new demand for domestic labor as women spent less time at home and more time at work. Immigrants contributed an estimated one-third of the total national GDP between 2000 to 2007 not only through the supply of much needed low-skilled labor, but through direct increases in consumption and demand for housing as well (Gálvez Iniesta, 2020).

¹As measured by percentage of the resident population.

From the 10-year period between 1998 to 2007 alone, the proportion of foreign-born persons registered as living in Spain increased over five-fold, from 1.6% to 10% of the total population. Estimates of the true percentage of foreign-born inhabitants, not just those registered in the municipal system, over the years 1970 and 2020 are shown in Figure 1 below.



Data source: UN – World Population Prospects

Figure 1: Counts of Spanish population by origin and percentage foreign-born (1970-2020)

The rapid influx of immigrants during this time presents a unique setting to study how initial settlement patterns of individuals, often determined with little understanding of local contexts and dictated by broader forces of the migratory network, can have an impact on the economic integration of recently arrived migrants in Spain.

2 DATA

2.1 *Encuesta nacional de inmigrantes* (ENI) from 2007

Primary data for this analysis comes from the Spanish National Institute of Statistics' (INE) National Survey of Immigrants (ENI). The survey population consists of 15,465 immigrant households living in Spain as of January 2007 that were randomly selected from predesignated pools of households with at least one individual registered as living there who met the following requirements: The individual 1) was born outside of Spain; 2) was 16 years or

older at time of surveying; 3) had established residency in Spain for at least one year prior to surveying and if not the case, had the intention of staying in Spain for at least a year; and 4) for those who had Spanish nationality from birth (but born outside of Spain), must have been at least 2-years-old at the time of arrival to Spain.

The sampling process was defined at the autonomous community level (NUTS-2) on differing population strata to target certain groups of nationalities of interest, particularly those from Ecuador, Morocco, Romania, Latin American countries excluding Ecuador, African countries excluding South Africa, Asian countries excluding Japan, North American excluding Mexico and Oceania, and the EU15 and European Economic Area excluding EU member countries. 87.37% of the desired sample size was achieved (15,465 of 17,700), with 57.62% of surveyed households coming from the primary targeted population and 42.38% coming from the supplementary targeted population. 144 countries of origin are captured in the final ENI sample, with Moroccans, Romanians, Ecuadorians, and Colombians being the most represented nationalities (12.0%, 8.6%, 8.5%, and 6.7%, respectively). All 50 Spanish provinces (NUTS-3) are represented in the sample, with Madrid, Barcelona, Illes Balears, and Murcia being the most popular provinces of residence (12.1%, 8.5%, 7.6%, and 7.1%, respectively).

When it comes to characterizing the cohorts of immigrants arriving to Spain, the ENI is advantageous in that its highly detailed questionnaire captures over 1,500 columns of data that identify the “demographic and social characteristics of those persons born abroad, as well as their migratory itineraries, work and residential history, family relations and relations with their country of origin and with the Spanish society” (Reher et al., 2008).² Missing information is, however, one downfall of such a broadly targeted and highly detailed survey. Mainly, not all who were surveyed in the ENI were employed at the time as the survey was not focused solely on measuring immigrants’ economic lives. Current income and employment information is available for about half of respondents (49.7%) while employment history is available for about two-thirds of respondents (65.7%). While other annual surveys such as the Annual Wage Structure Survey and Continuous Household Survey also provide micro-level information on Spanish residents including those foreign-born, neither is designed to measure the social, economic, and demographic structure of *immigrants* exclusively, nor do they report explicit data on job-related earnings or detailed data on migratory and settlement history. In this sense, the ENI provides an incomparable level of detail that can be used to understand better the potential effects of co-location of immigrants on individual and group-level economic outcomes.³

The broad definition of “immigrant” employed in the ENI begs the need of differentiating

²Link to the raw data, dictionaries, survey methodology, and questionnaire used to collect responses can be found at https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177005&menu=resultados&idp=1254735573002#!tabs-1254736195389

³More detailed information about the ENI’s surveying methods, descriptive statistics, and diagnostics can be found in Reher & Requena (2009).

between different types of foreign residents and their underlying motivations for coming to live in Spain. After all, there are most likely underlying differences between those who decide to locate themselves where they expect good job prospects to be versus those who, say, are looking for a sunny beach to retire to. Migratory motives are provided in the ENI data, so I choose to filter to the 8,604 individuals I deem “economic immigrants”, or those who have come to Spain for 1) lack of work in home country, 2) to look for better employment, or 3) for higher quality of life so long as they were not from a country classified as “developed”.⁴

⁵ Making the distinction between economic and non-economic immigrants helps ensure that any unobserved factors that could simultaneously affect settlement patterns *and* economic outcomes are at least arguably more similar within these groups than across them, and therefore easier to address. To check the internal validity of this measure of motivation, I conduct a two-sample t-test and find that the proportion of those currently working at the time of surveying is significantly greater for the sample of economic immigrants than that for sample of the non-economic immigrants (Table 1).

⁴Among the other options to choose from were for retirement, political or religious reasons, family reunification, formative or educational reasons, climate, or for a temporal stay while transitioning to another country. An individual could check off two or more motives if relevant. Therefore, to ensure that a retiree who answered as their reasons for moving to Spain as both for retirement *and* for lack of work in home country does not get labelled as an economic immigrant, only those who checked off just one or both of the two aforementioned motives qualify.

⁵Developed countries are defined as those that had a recorded GDP per capita that was at least 75% or more of that of Spain in 2007.

Table 1: Comparison of Economic and Non-Economic Immigrants in the ENI

| Variable | Economic Immigrants | | | Non-economic Immigrants | | | Test of Equal Props. |
|---|---------------------|-------|---------|-------------------------|-------|---------|----------------------|
| Value | Count | Total | Percent | Count | Total | Percent | p-value |
| Currently working? | | | | | | | |
| No | 2227 | 8596 | 25.91 | 3379 | 6850 | 49.33 | 1.79e-198 |
| Yes | 6369 | 8596 | 74.09 | 3471 | 6850 | 50.67 | 1.79e-198 |
| Gender | | | | | | | |
| Male | 4291 | 8604 | 49.87 | 2673 | 6861 | 38.96 | 9.67e-42 |
| Female | 4313 | 8604 | 50.13 | 4188 | 6861 | 61.04 | 9.67e-42 |
| Level of studies | | | | | | | |
| No secondary | 2531 | 8594 | 29.45 | 1661 | 6854 | 24.23 | 4.97e-13 |
| Secondary (1st cycle) | 1336 | 8594 | 15.55 | 1294 | 6854 | 18.88 | 4.88e-08 |
| Secondary (2nd cycle) | 3295 | 8594 | 38.34 | 2068 | 6854 | 30.17 | 3.76e-26 |
| Tertiary (1st cycle) | 1330 | 8594 | 15.48 | 1658 | 6854 | 24.19 | 3.84e-42 |
| Tertiary (2nd cycle) | 102 | 8594 | 1.19 | 173 | 6854 | 2.52 | 6.28e-10 |
| Country classification | | | | | | | |
| Andean countries | 2487 | 8604 | 28.91 | 721 | 6861 | 10.51 | 1.18e-172 |
| Asia and Oceania | 263 | 8604 | 3.06 | 107 | 6861 | 1.56 | 1.97e-09 |
| Caribbean and others | 5 | 8604 | 0.06 | 5 | 6861 | 0.07 | 0.968 |
| Developed countries (>75% GDPpc of Spain) | 717 | 8604 | 8.33 | 3181 | 6861 | 46.36 | 0 |
| Eastern Europe, Balkans, other European nationalities | 1921 | 8604 | 22.33 | 473 | 6861 | 6.89 | 7.1e-153 |
| Northern Africa & Middle East | 1248 | 8604 | 14.5 | 888 | 6861 | 12.94 | 0.00554 |
| Other Latin American countries | 1617 | 8604 | 18.79 | 1343 | 6861 | 19.57 | 0.228 |
| Sub-Saharan Africa (excl. S. Africa) | 346 | 8604 | 4.02 | 143 | 6861 | 2.08 | 1.1e-11 |
| Top sectors of work | | | | | | | |
| Agriculture | 464 | 6369 | 7.29 | 99 | 3471 | 2.85 | 2.22e-19 |
| Business, commerce | 580 | 6369 | 9.11 | 498 | 3471 | 14.35 | 2.38e-15 |
| Construction | 1393 | 6369 | 21.87 | 330 | 3471 | 9.51 | 1.84e-53 |
| Health, social, and community services | 416 | 6369 | 6.53 | 407 | 3471 | 11.73 | 8.39e-19 |
| Hospitality | 835 | 6369 | 13.11 | 399 | 3471 | 11.5 | 0.0226 |
| Household activities | 1035 | 6369 | 16.25 | 210 | 3471 | 6.05 | 1.02e-47 |
| Manufacturing | 718 | 6369 | 11.27 | 408 | 3471 | 11.75 | 0.494 |
| Real estate; rental, business services | 448 | 6369 | 7.03 | 414 | 3471 | 11.93 | 3.17e-16 |
| Unspecified | 5 | 6369 | 0.08 | 7 | 3471 | 0.2 | 0.171 |
| Autonomous Community | | | | | | | |
| Andalucía | 599 | 8604 | 6.96 | 634 | 6861 | 9.24 | 2.37e-07 |
| Aragón | 395 | 8604 | 4.59 | 194 | 6861 | 2.83 | 1.61e-08 |
| Asturias (Principado de) | 200 | 8604 | 2.32 | 303 | 6861 | 4.42 | 4.49e-13 |
| Balears (Illes) | 626 | 8604 | 7.28 | 556 | 6861 | 8.1 | 0.0581 |
| Canarias | 318 | 8604 | 3.7 | 358 | 6861 | 5.22 | 5.13e-06 |
| Cantabria | 227 | 8604 | 2.64 | 196 | 6861 | 2.86 | 0.437 |
| Castilla y León | 355 | 8604 | 4.13 | 282 | 6861 | 4.11 | 0.993 |
| Castilla-La Mancha | 460 | 8604 | 5.35 | 151 | 6861 | 2.2 | 2.94e-23 |
| Catalunya | 1102 | 8604 | 12.81 | 718 | 6861 | 10.46 | 7.92e-06 |
| Ceuta | 29 | 8604 | 0.34 | 75 | 6861 | 1.09 | 1.95e-08 |
| Comunidad Valenciana | 682 | 8604 | 7.93 | 769 | 6861 | 11.21 | 4.33e-12 |
| Extremadura | 197 | 8604 | 2.29 | 206 | 6861 | 3 | 0.00665 |
| Galicia | 245 | 8604 | 2.85 | 347 | 6861 | 5.06 | 1.5e-12 |
| Madrid (Comunidad de) | 1152 | 8604 | 13.39 | 725 | 6861 | 10.57 | 1.07e-07 |
| Melilla | 63 | 8604 | 0.73 | 93 | 6861 | 1.36 | 0.000162 |
| Murcia(Región de) | 694 | 8604 | 8.07 | 401 | 6861 | 5.84 | 1.04e-07 |
| Navarra(Comunidad Foral de) | 625 | 8604 | 7.26 | 439 | 6861 | 6.4 | 0.0374 |
| País Vasco | 295 | 8604 | 3.43 | 272 | 6861 | 3.96 | 0.0857 |
| Rioja (La) | 340 | 8604 | 3.95 | 142 | 6861 | 2.07 | 3.04e-11 |

I also find that there are significant differences in the gender, educational level, country of origin, and sectoral compositions of the two sample populations. Whereas non-economic immigrants are more likely to be female due to the fact that the majority of women in this group come for family reunification reasons, economic immigrants are more likely to be newer arrivals from developing countries with shorter migratory histories to Spain such as Romania, Ecuador, Colombia, and Bolivia. While out of all those surveyed, many reside in the metropolitan regions of Madrid and Catalunya, there is proportionately more economic immigrants living in Murcia, Navarra, and Aragón than non-economic immigrants. Economic immigrants also tend to be younger on average, have lower rates of tertiary educational attainment, and are more concentrated in the agricultural, construction, and household sectors.⁶

2.1.1 Response Variable: Log of hourly wage

Working respondents were asked to report how much money they receive a month for their jobs, including the monthly proportion of any regularly-earned outstanding payments, as well as how many hours a week they work habitually at their jobs. Of the 6,369 economic immigrants that reported their current laboral situation as “working” (74% of all economic immigrants), 5,093 reported their monthly earnings with a real value as well as their weekly worked hours (80% of working economic immigrants, 59% of all economic immigrants).⁷ Estimated hourly earnings are calculated by dividing total monthly earnings by the estimated total hours worked habitually a month.⁸

Table 2 shows the average reported hourly wage and sample size for each group from country c living in province p at the time of surveying. Countries of origin and provinces displayed are those with the most representation among the entirety of the population sample of economic immigrants with reported earnings who arrived in Spain in 1998 or later.⁹ While average wage levels differ at both the country-of-origin level and the province level, there are also differences in wage level within groups. For example, while Peruvians earn higher wages than the average economic immigrant, they earn about 12% less than the province-wide average in Zaragoza but 13% more than the province-wide average in Barcelona. Romanians, on the other hand, make about 3% more in Zaragoza but 13% less in Barcelona. Of course, these within-group differences could very well be due to underlying systematic differences in age, education level, socioeconomic background, or sector of work. The empirical analysis shows,

⁶See Table 7 in Appendix for distributions of age and years in Spain.

⁷Reported monthly earnings were not adjusted for provincial differences in the Consumer Price Index as systematic variation in wages due to location will be accounted for with geographical controls.

⁸To estimate total hours worked habitually a month, total hours worked habitually a week are multiplied by 4.

⁹See Section 2.2.1 for rationale behind this restriction.

however, that these within-group regional differences still remain after controlling for various observable demographic, human capital, and work-related factors.

Table 2: Mean hourly wages for top national origin groups in selected provinces

| | Argentina | Bolivia | Colombia | Ecuador | Morocco | Peru | Romania | Total |
|-------------------------|------------|------------|-----------|------------|------------|------------|------------|-------------|
| Baleares (Illes) | 7.74 (45) | 4.54 (16) | 5.86 (24) | 6.07 (36) | 6.36 (30) | 6.95 (6) | 8.25 (19) | 7.13 (282) |
| Barcelona | 8.04 (39) | 5.87 (42) | 6.4 (41) | 6.43 (71) | 6.3 (43) | 8.02 (31) | 6.18 (20) | 7.08 (436) |
| Madrid | 9.57 (20) | 4.27 (43) | 6.14 (65) | 5.6 (130) | 5.33 (28) | 6.03 (46) | 5.71 (138) | 5.97 (640) |
| Murcia | 6.96 (10) | 5.03 (45) | 5.49 (19) | 5.82 (163) | 5.67 (91) | 5 (1) | 5.31 (28) | 5.69 (459) |
| Navarra | 6.95 (11) | 4.02 (20) | 6.4 (43) | 6.6 (111) | 6.24 (28) | 6.82 (11) | 7.56 (16) | 6.61 (380) |
| Rioja (La) | 12.5 (1) | 5.34 (16) | 5.77 (18) | 5.49 (19) | 5.03 (8) | 5.92 (5) | 5.82 (69) | 5.9 (179) |
| Valencia | 5.59 (8) | 3.59 (12) | 5.3 (25) | 5.24 (27) | 5.51 (9) | 7.44 (2) | 5.65 (26) | 5.47 (182) |
| Zaragoza | NA | 4.61 (5) | 6.35 (5) | 6.15 (27) | 5.67 (16) | 5.26 (6) | 6.17 (62) | 6 (167) |
| Total | 7.43 (225) | 4.73 (250) | 6 (401) | 5.96 (699) | 5.82 (401) | 6.86 (145) | 5.87 (690) | 6.17 (4135) |

¹ Sample size of each group shown in parentheses.

² Column-wise means include all provinces and row-wise means include all countries of origin, not just those displayed.

2.2 *Estadística del Padrón continuo (Padrón data) from 1998 - 2007*

Aggregate population data by year and Spanish province is taken from the municipal registry database (*Padrón municipal*) in order to map province-level yearly residential population data to immigrants in the ENI by country of origin.¹⁰ Registry data is available at the province-level (NUTS-3) and at the municipal-level. However, only province-level data will be considered since 1) municipal-level data is only available from 2003 onwards and 2) linking municipal-level population data to individuals in the ENI is not currently possible due to data protection reasons.^{11 12}

Registering in the local municipality system is obligatory for *anyone* who resides in Spain. There are incentives to do so for Spanish and EU nationals as it provides access to public medical care and education. Residency authorizations for foreigners are contingent upon being *empadronado*. Even undocumented immigrants have motives to register themselves as the regularization mechanism of *arraigo*, established in 2004, grants legal status to foreigners

¹⁰Of the 112 countries of origin represented in the study population of economic immigrants from the ENI, the *Padrón* data is summarized by exact nationality for 97. Those from Azerbaiyán, Uzbekistán, Mozambique, Tanzania, Chad, Zimbabwe, Trinidad and Tobago, Suriname, Puerto Rico, and Sri Lanka, represented by a total of 16 respondents, are categorized in aggregated groups corresponding to those in the *Padrón* data.

¹¹Information on municipality of residence in Spain and history of movements between municipalities is available in the ENI, but fictitious names of the municipalities are assigned for data protection reasons in accordance with the *Secreto Estadístico* of Spain. A fee of 367,84€ must be paid in order to access this data.

¹²Link to the aggregated data files with registry information counted by year, province, and nationality can be found at <https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p04/provi/10/&file=0ccaa002.px&L=0>

who, conditional on various other requirements, can prove that they have been living in Spain for at least two years (Gálvez Iniesta, 2020). Thus, the *Padrón* data is advantageous for the purposes of this study as it provides accurate residential data on the broad range of foreign-born populations living in Spain as of 2007. However, it is these same policy changes and the incentives they create for registering in the system that could introduce great measurement bias into the data, an issue I will turn back to in Section 4.2.1.

2.2.1 Explanatory Variable: Co-ethnic density

A measure of co-ethnic density assigned to each individual is defined as the proportion of the province’s total resident population who are of the same national origin and gender. Co-ethnic matching is done by gender with the understanding that within the segregated labor markets of natives and immigrants, there is further segregation on the lines of gender. In the ENI population, the top-represented sectors of work are highly segregated. For example, the construction sector is 98.7% male and 1.3% female, household activities are 2.6% male, 97.4% female, hospitality 31.0% male, 69.0% female, and manufacturing 71.7% male, 28.3% female. Thus, if co-ethnic density has an effect on the economic lives of immigrants, that effect will be best captured through co-ethnic and co-gender networks.

The co-ethnic density index can thus be formalized as such:

$$\text{Co-Ethnic Density} = \frac{N_{c,g,p}}{N_{g,p}} \quad (1)$$

where $N_{c,g,p}$ represents the number of individuals born in foreign country c of gender g living in Spain in province p as of the 1st of January of the year in which the individual of reference reported having begun living in that province. Since *Padrón* data stratified by province and nationality is only available from 1998 onwards, the study population will be restricted to those who arrived to province p in 1998 or later.¹³ $N_{g,p}$ therefore represents the *total* number of residents of same gender g living in province p in that same year, foreign-born and natives alike.¹⁴

¹³Any ENI respondents residing in the autonomous cities of Ceuta and Melilla are excluded from the sample.

¹⁴There are 24 individuals in the study population who live in provinces where no other co-ethnics are registered as living in in the year the individual started living there. These are principally those who come from small countries with little pre-established migratory history to said provinces. Since the case of no other co-ethnic residents living in such a geographically large area is improbable and since I know that at least the one ENI respondent was reportedly living in the province in said year, I impute the missing numerator $N_{c,g,p}$ with a value of one.

Table 3: Mean co-ethnic density per 1,000 residents for top 7 most represented national origin groups in selected provinces

| | Argentina | Bolivia | Colombia | Ecuador | Morocco | Peru | Romania | Total |
|-------------------------|------------|------------|------------|--------------|--------------|------------|-------------|--------------|
| Baleares (Illes) | 5.94 (3.8) | 2.09 (1.5) | 5.62 (3.3) | 7.14 (5) | 15.39 (6) | 0.94 (0.5) | 1.68 (1.6) | 5.41 (5.9) |
| Barcelona | 3.27 (1.7) | 3.52 (2.2) | 3.56 (2.2) | 7.91 (5.8) | 16.95 (6.7) | 3.55 (1.3) | 1.77 (1.1) | 4.86 (6) |
| Madrid | 2.37 (1.1) | 4.09 (2) | 7.45 (4.9) | 13.75 (11.5) | 9.88 (3.4) | 5.39 (1.7) | 10.4 (5.9) | 7.42 (7.8) |
| Murcia | 1.19 (0.4) | 3.75 (2.9) | 2.13 (1.8) | 20.81 (16.3) | 28.36 (15.4) | 0.14 (NA) | 2.24 (1.2) | 13.73 (16.2) |
| Navarra | 1.5 (0.4) | 2.08 (1.4) | 5.16 (3.2) | 12.65 (8.8) | 7.15 (4) | 1.63 (0.6) | 2.23 (1.4) | 5.82 (7) |
| Rioja (La) | 2.39 (NA) | 2.78 (2.1) | 5.89 (4) | 4.78 (3.7) | 16.89 (9.9) | 0.8 (0.4) | 9.42 (6.9) | 6.95 (6.9) |
| Valencia | 1.74 (1.3) | 3.61 (2) | 3.45 (2.5) | 4.5 (4.5) | 5.73 (2.6) | 0.51 (0.1) | 5.7 (3.8) | 3.05 (3.2) |
| Zaragoza | NA | 0.32 (0.2) | 1.42 (0.4) | 4.08 (4.3) | 4.78 (3.2) | 0.56 (0.2) | 12.53 (8.2) | 5.81 (7.4) |
| Total | 3.49 (3) | 2.9 (2.4) | 4.67 (3.8) | 11.37 (11.8) | 15.49 (13.3) | 2.92 (2.3) | 9.67 (10.7) | 6.6 (9.4) |

¹ Standard deviations shown in parentheses.

² Column-wise means include all provinces and row-wise means include all countries of origin, not just those displayed.

The column-wise variability in co-ethnic density we see in Table 3 shows that immigrants originating from the same country choose to settle in different provinces at different times, but that these geographical settlement preferences are not necessarily similar across groups (rows), providing evidence that co-ethnic networks play a role in the settlement patterns of immigrants in Spain. The empirical analysis that follows investigates how these networks may continue to play a role in the economic lives of immigrants once settled.

3 METHODOLOGY

To test whether co-ethnic density, $COED$, has an effect on the log of hourly wage, $\log WAGE$, of immigrant i , I will utilize the following theoretical model:

$$\log(WAGE_i) = \beta_0 + \beta_1 COED_i + \rho X_i + \zeta SECT_i + \gamma COUNTRY_i + \psi REG_i + \epsilon_i \quad (2)$$

where X_i denotes a vector of individual-level covariates capturing level of education attained and the number of years since migration as well as demographic variables for age, the square

of age and gender of the individual.¹⁵ ¹⁶ ¹⁷ $SECT_i$ denotes a control for sector of work to unconfound any co-ethnic residential clustering that may be a result of sectoral regionalization such as in the agricultural or manufacturing industries, the former localized in the less dense southern provinces in Andalucía and Murcia while the latter in the more concentrated regions of the north.¹⁸ Finally, $COUNTRY_i$ and REG_i denote vectors of dummy variables indicating country of origin and region of residence, respectively. Region-level fixed effects allows for differing wage levels between geographical regions that affect all groups within the region similarly due to general differences in cost-of-living or labor market conditions, a prominent feature of the Spanish system (see Maza & Villaverde (2009) and Liu (2018) for the relevant discussion). Country-of-origin fixed effects, on the other hand, controls for factors that arguably affect all co-ethnic individuals similarly when it comes to economic integration and assimilation in a foreign country, such as language, culture, or institutions. Regions are stratified within provinces by population, and while this does not account for the geographical size of the municipality, the distinction between urban and rural regions of each province generally correspond to the population stratifications utilized, this relationship visualized in Figure 3 of the Appendix where one can see high population count municipalities clustered around the major metropolitan areas. Finally, to avoid singularities, I restrict the population to the top 43 most represented countries of origin and top 111 most represented regions.¹⁹

Since estimating the effect of co-ethnic density on wages is only possible to do for those work-

¹⁵Level of completed education is reported on the ENI as one of the following categories: “No studies”, “Some studies”, “Primary studies (incomplete)”, “Primary studies (complete)”, “Secondary studies (1st cycle)”, “Secondary studies (2nd cycle)”, “Tertiary studies (1st cycle)”, and “Tertiary studies (2nd cycle)”. While the use of *years* of completed education is common when studying the return to education in the United States context (see Card (1999) for the relevant discussion), I opt to maintain categorical factors representing completed levels of education in the following hierarchical order: “No secondary studies”, “Secondary studies (1st cycle)”, “Secondary studies (2nd cycle)”, “Tertiary studies (1st cycle)”, and “Tertiary studies (2nd cycle)”. Public education in Spain is free and obligatory up until the age of 16, the year in which an individual can stop studying or continue his secondary studies in the 2nd cycle (*Bachillerato* or *formación profesional de Grado Medio*). Therefore, it is unlikely that differences in schooling level for those who do not have any secondary education would have a substantially different impact on wages in the Spanish context.

¹⁶According to the seminal work of Chiswick (1978), when measuring the earnings of immigrants individual-level human capital can be proxied for by time spent in the destination country as one goes assimilating to the new culture, job market, norms, language, and general way of life in which he finds himself.

¹⁷Information on birth year and arrival year was missing for 28 individuals. Values for age and number of years since migration were imputed for these individuals via conditional mean imputation, regressing them individually on all other independent variables in the outcome model for the set of observations with complete information and using the model to predict the missing values. See Appendix for specifications of the imputation model.

¹⁸A series of bilateral agreements reached in the early 2000s expanded work visa quotas in the construction sector coming from some of the largest immigrant exporting countries at the time like Ecuador, Colombia, Dominican Republic, and Morocco highlighting how regional sectors can influence observed co-ethnic clustering (Hooper, 2019).

¹⁹For each grouping, $n > 10$.

ing and hence generating income, I use Heckman’s sample selection model to test whether there are any systematic differences between the populations of economic immigrants who reported their wages versus those who did not, understanding that these differences could bias my estimates (Heckman, 1979). For example, immigrant women whose cultures restrict them from working outside the home, often those that also restrict access to education and other basic rights for women, may be more likely to reside in highly dense co-ethnic areas. These women, if they could work, would on average make lower wages, so by not observing them the estimates for the effect of co-ethnic density on wages may be biased upwards. Comparing the two populations in Table 8 of the Appendix, we see that women are less likely to have been observed as working than men are, particularly for those women who are married with children living in the same household. We also see that those from Andean countries and Eastern Europe, the Balkans, and other European nationalities are more likely to be working whereas those from developed countries and northern Africa & the Middle East are less likely to be working. Given these results, I specify the selection equation such that the probability of having wage data depends on whether one is female or not, what country classification his or her country of origin belongs to (classifications shown in Table 9), and whether or not he or she is a married woman with children in the house. Regressing a Probit model on the selection equation, I calculate the inverse Mills ratio, $\hat{\lambda}$, and estimate the wage model with $\hat{\lambda}$ as an added regressor. I find correlation between the error terms of the selection and outcome equations in all specifications of the wage model such that I reject $H_0 : \hat{\lambda} = 0$ and claim that there is significant reason to believe that there is a problem of sample selection when estimating my wage model.²⁰ Therefore, all results shown in Section 4 are based on the Heckman selection model.

When analyzing the effect of co-ethnic density on the wages of immigrant i , the coefficient of interest, β_1 , represents the difference in the log of hourly wage, or the percentage change in hourly wage, if that same person were to have settled in a province with one percentage point more of the population made up of co-ethnics. Due to the lack of theoretical understanding of the mechanisms at work and the contradicting empirical evidence about the directionality of the effect as discussed previously, I now turn to the data to see if there is any independent effect of co-ethnic density on immigrant wages.

²⁰Significance level determined at $\alpha \leq 0.05$

4 RESULTS

4.1 Estimating Effect of Co-Ethnic Density via OLS

Table 4 presents the results from the Heckman sample selection wage model using various specifications of Equation 2 to estimate the effect of co-ethnic density, β_1 on hourly earnings of economic immigrants. Models 1 through 8 show the results of the adjusted OLS regressions, progressively controlling for more covariates with each iteration of the model. I expect substantial multicollinearity to exist between my explanatory variable of interest and the other covariates in the model given the fact that residential location, in the case of my sample, is not determined randomly. Indeed, the initial negative effect of co-ethnic concentration seen in Model 1, representing about a 1.1% decrease in monthly earnings for every one percentage point increase in co-ethnic density, increases in magnitude by almost three when controlling for gender of the individual indicating that males, who on average make significantly more than their female counterparts, also tend to settle in more relatively dense co-ethnic provinces. Those who settle in more relatively dense co-ethnic provinces also tend to be newer immigrants with lower education level who make less on average than their more established and educated counterparts, as seen by the change in the estimated effect of co-ethnic density when controlling for years in Spain in Model 3 and educational achievement in Model 4. However, the independent effect of co-ethnic density on wages remains negative and significant even after the addition of controls for age, the square of age, and sector of work. The additional fixed effect for country of origin (Model 7) shifts the co-ethnic density effect estimate slightly upwards, indicating that those from certain countries of origin who have lower earning levels as a whole tend to settle in more concentrated provinces. Finally, the addition of the fixed effect for region of residence in the fully specified model (Model 8) does not change substantially the results. While insignificant at $\alpha = 0.1$, we do reject the null hypothesis that $COED = 0$ at $\alpha = 0.14$

To interpret concretely what the results from Model 8 suggest, we may understand the effect of co-ethnic density on wages as such: the wage of individual i in a parallel universe would be about 1% greater if the only thing that had changed about her situation was that she had decided to settle in a province with one percentage point less of the residents hailing from her home country. This effect is independent of region, which is perhaps most surprising given that I expect certain metropolitan areas to attract more immigrants to them due to the fact that they have higher wage levels and better employment opportunities. If this is indeed the case, it suggests that the true effect of co-ethnic density may be even more negative.

Table 4: Effect of co-ethnic density on the log of hourly wage

| | <i>Dependent variable:</i> | | | | | | | |
|-------------------------|----------------------------|----------------------|----------------------|----------------------|------------------------|------------------------|------------------------|------------------------|
| | Log of hourly wage | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| COED | −0.012* (0.007) | −0.030*** (0.006) | −0.023*** (0.006) | −0.016** (0.006) | −0.015** (0.006) | −0.014** (0.006) | −0.009 (0.007) | −0.010 (0.007) |
| Female | | −0.303*** (0.016) | −0.280*** (0.014) | −0.295*** (0.015) | −0.290*** (0.015) | −0.191*** (0.018) | −0.221*** (0.029) | −0.212*** (0.028) |
| Years in Spain | | | 0.016*** (0.002) | 0.016*** (0.002) | 0.016*** (0.002) | 0.013*** (0.002) | 0.013*** (0.002) | 0.013*** (0.002) |
| Secondary (1st cycle) | | | | 0.034* (0.019) | 0.031* (0.019) | 0.018 (0.018) | −0.011 (0.018) | −0.007 (0.018) |
| Secondary (2nd cycle) | | | | 0.071*** (0.014) | 0.067*** (0.014) | 0.051*** (0.014) | 0.024 (0.014) | 0.022 (0.014) |
| Tertiary (1st cycle) | | | | 0.191*** (0.019) | 0.185*** (0.019) | 0.148*** (0.019) | 0.097*** (0.019) | 0.101*** (0.019) |
| Tertiary (2nd cycle) | | | | 0.405*** (0.059) | 0.399*** (0.059) | 0.317*** (0.058) | 0.217*** (0.058) | 0.200*** (0.057) |
| Age | | | | | 0.015*** (0.004) | 0.013*** (0.004) | 0.015*** (0.004) | 0.015*** (0.004) |
| Age squared | | | | | −0.0002*** (0.0001) | −0.0002*** (0.0001) | −0.0002*** (0.0001) | −0.0002*** (0.0001) |
| Sector controls | No | No | No | No | No | Yes | Yes | Yes |
| Origin country controls | No | No | No | No | No | No | Yes | Yes |
| Region controls | No | No | No | No | No | No | No | Yes |
| Observations | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 |

Note:

*p<0.1; **p<0.05; ***p<0.01

All results shown are based on the Heckman selection model.

4.2 Estimating Effect of Co-Ethnic Density (Absolute and Relative) via 2SLS

The results presented above suggest that the overall effect of residing in a highly dense co-ethnic area is bad for immigrant wages. However, it does not necessarily shed light on the mechanisms behind what may very well be the net impact of various positive and negative forces. The theory that ethnic enclaves, especially linguistic ones, can reduce rates of interaction with the native population as well as the need to improve foreign language ability or assimilate other foreign-acquired human capital has found empirical support: Chiswick & Miller (2005) show that ethnic clustering in the United States is associated with lower earnings amongst male immigrants from non-English speaking countries along with lower English attainment, while Kanas, Chiswick, Van Der Lippe, & Van Tubergen (2012) provide evidence that residential co-ethnic concentration in Germany has a weak negative effect on the economic returns of education achieved in the host country.

However, research on immigrant enclaves has spurred heated debate on how to operationalize

such a complex phenomenon (Portes & Shafer, 2007). Therefore, I shall take extra care to emphasize that I am *not* measuring the effect of ethnic enclaves on economic outcomes of immigrants, and indeed the explanatory variable of interest I have employed thus far is very far from being able to identify these highly localized and contextualized sociocultural and economic relationships. While my measure of co-ethnic density as defined above may be capturing the effect of networks that arise from living in relatively dense co-ethnic areas, it could also represent, and perhaps more intuitively represents, the magnitude of immigrant inflows given that the immigrant share of the total population during the study period was growing at a very fast rate at the time (Figure 1). If this is the case, then it would make sense that wages for immigrants are depressed in areas where there are greater stocks of immigrant labor since greater job competition arises under the assumption that arriving immigrant labor is a close substitute for that of already arrived ones (Borjas, Freeman, Katz, DiNardo, & Abowd, 1997). Since upward mobility is difficult for the otherwise qualified immigrant who faces legal, institutional, and language barriers to accessing the broader job market, more and more get filtered into the more easily accessible low-paying secondary sectors where less specialization and human capital assimilation is required, the same sectors that were in search of cheap labor during this sustained period of prosperity and growth in the Spanish economy. In fact, this is the theory employed in other empirical studies that have shown high rates of over-education in certain immigrant populations (Joona, Gupta, & Wadensjö, 2014) (Hou, Lu, & Schimmele, 2021).

With this argument in mind, I opt to include a measure of *relative co-ethnic density* into Equation 2, which adjusts the co-ethnic density index from Equation 1, now referred to as the *absolute co-ethnic density*, for each individual from country c of gender g living in province p by the absolute co-ethnic density index of all individuals from country c of gender g living in Spain such that:

$$\text{Relative Co-Ethnic Density} = \frac{\frac{N_{c,g,p}}{N_{g,p}}}{\frac{N_{c,g}}{N_g}} \quad (3)$$

where $N_{c,g}$ represents the number of individuals born in foreign country c of gender g living in Spain in the year of the individual of reference's arrival in province p and N_g represents the number of all individuals of gender g living in Spain in that same year. In this way, even if the overall counts for some of the cohorts are small, the relative co-ethnic density index will be more able to capture differences in network *strength*, whereas the absolute measure of co-ethnic density is better at capturing network *magnitude*.

To make this idea more concrete, in 2004, immigrants from Cabo Verde made up only about 0.006% of the total Spanish resident population. That means that for every 100,000 residents in Spain there were 6 from Cabo Verde. However, there were nearly 100 Cabo

Verdeans for every 100,000 residents registered in León at the time. That means that Cabo Verdeans were over sixteen times more concentrated in León than if they were to perfectly disperse themselves throughout any of the other provinces. The size of this effect for what is a relatively low-represented immigrant group in absolute terms suggests a specific quality about the migratory networks that were leveraged in the settlement patterns of Cabo Verdeans in Spain, and I hypothesize that these networks may continue having an effect on the economic lives of immigrants once settled.

Figure 2 illustrates the differences between the two indices for selected countries of origin. While the absolute measure of co-ethnic density index is able to capture patterns of residential clustering for those countries of origin with longer or more intense migratory histories like Morocco or Romania, it is only with the relative measure of co-ethnic density that these patterns emerge for countries with lower magnitude or more recent migratory flows like from Cabo Verde or Venezuela.

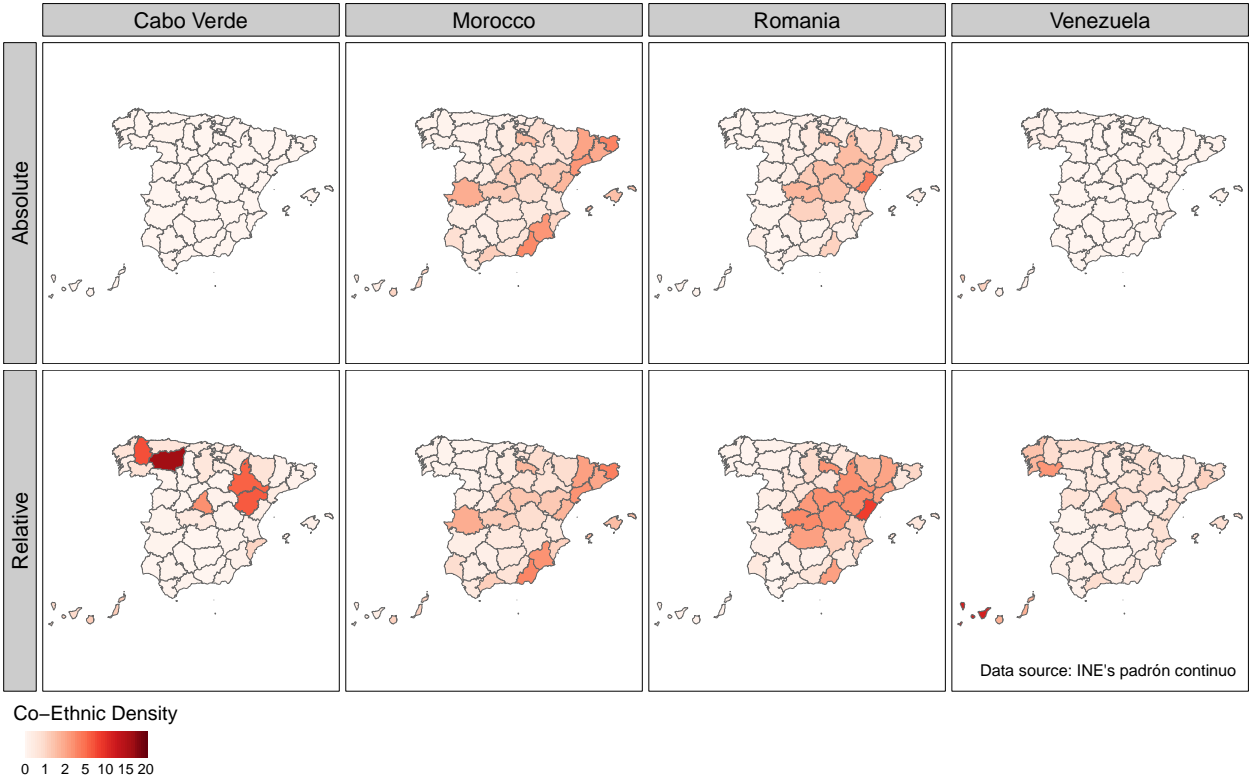


Figure 2: Absolute and Relative Co-Ethnic Density Measures for Select Countries of Origin (2004)

When estimating the relative co-ethnic density index on wages, I take the log of the index so that any negative value represents province co-ethnic densities that are less than the national density, which we may interpret as representing weaker networks, and positive values represent those with relatively more co-ethnic concentrations than the national average,

representing stronger networks. I also log the measure of absolute co-ethnic density to achieve more variance in the distribution at lower domain values and control for this such that my new wage equation looks as such:

$$\log(WAGE_i) = \beta_0 + \beta_1 \log(absCOED_i) + \beta_2 \log(relCOED_i) + \rho X_i + \zeta SECT_i + \gamma COUNTRY_i + \psi REG_i + \epsilon_i \quad (4)$$

We can see in Table 5 that the estimated effect for the log of relative co-ethnic density throughout all specifications of the model is positive and quite significant. Remembering that in this analysis absolute co-ethnic density is logged, we still see the same general pattern of a negative and significant effect as we saw in Table 4. The patterns of how the effect changes with each added covariate are also generally the same. It is interesting to note, however, that when controlling for country of origin in Model 7, the effect of absolute density becomes more negative unlike in Model 7 of Table 4 where it becomes more positive. This positive effect instead appears to “move” to the relative density effect. The shift from Model 6 to Model 7 in Table 4 indicates that those from certain countries of origin that have lower baseline wages are also those that have high measures of absolute density, or, as discussed before, greater overall magnitude of co-ethnics living there. However, when we control for differences in relative density, or the likelihood of settling in one area over another *within* ethnic groups, we see the negative effect of the former become more negative while the positive effect of the latter become more positive. In this way, by controlling for both absolute and relative ethnic density, the “stock” versus “strength” effect of migrant networks disentangle themselves. These results suggest that for a single immigrant it is more economically advantageous to live in a province with low absolute but high relative co-ethnic density, or more plainly put, within a strong but small co-ethnic network.

Table 5: Effect of log of co-ethnic density (absolute and relative) on log of hourly wage

| | <i>Dependent variable:</i> | | | | | | | |
|-------------------------|----------------------------|----------------------|----------------------|----------------------|------------------------|------------------------|------------------------|------------------------|
| | Log of hourly wage | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| logCOED (absolute) | −0.023*** (0.005) | −0.027*** (0.004) | −0.019*** (0.004) | −0.014*** (0.004) | −0.014*** (0.004) | −0.012*** (0.004) | −0.019*** (0.006) | −0.020*** (0.006) |
| logCOED (relative) | 0.027*** (0.008) | 0.024*** (0.008) | 0.016** (0.008) | 0.015** (0.008) | 0.015** (0.008) | 0.013* (0.007) | 0.022*** (0.008) | 0.015 (0.009) |
| Female | | −0.297*** (0.016) | −0.276*** (0.015) | −0.292*** (0.015) | −0.287*** (0.015) | −0.190*** (0.018) | −0.216*** (0.028) | −0.206*** (0.028) |
| Years in Spain | | | 0.016*** (0.002) | 0.016*** (0.002) | 0.016*** (0.002) | 0.013*** (0.002) | 0.012*** (0.002) | 0.012*** (0.002) |
| Secondary (1st cycle) | | | | 0.035* (0.019) | 0.032* (0.019) | 0.018 (0.018) | −0.010 (0.018) | −0.006 (0.018) |
| Secondary (2nd cycle) | | | | 0.072*** (0.014) | 0.067*** (0.014) | 0.051*** (0.014) | 0.024* (0.014) | 0.022 (0.014) |
| Tertiary (1st cycle) | | | | 0.191*** (0.019) | 0.186*** (0.019) | 0.148*** (0.019) | 0.098*** (0.019) | 0.102*** (0.019) |
| Tertiary (2nd cycle) | | | | 0.405*** (0.059) | 0.400*** (0.059) | 0.317*** (0.058) | 0.216*** (0.058) | 0.198*** (0.057) |
| Age | | | | | 0.014*** (0.004) | 0.012*** (0.004) | 0.015*** (0.004) | 0.014*** (0.004) |
| Age squared | | | | | −0.0002*** (0.0001) | −0.0002*** (0.0001) | −0.0002*** (0.0001) | −0.0002*** (0.0001) |
| Sector controls | No | No | No | No | No | Yes | Yes | Yes |
| Origin country controls | No | No | No | No | No | No | Yes | Yes |
| Region controls | No | No | No | No | No | No | No | Yes |
| Observations | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 | 6,525 |

Note:

*p<0.1; **p<0.05; ***p<0.01

All results shown are based on the Heckman selection model.

4.2.1 Endogeneity

Given its spatial nature, co-ethnic density as defined above would be a more accurate measure of network effects if it could be measured at a more disaggregate level. However, as mentioned previously, information on the municipality of residence is anonymized on the ENI data making it impossible to link respondents to the municipality-level *Padrón* data. While maintaining my measure of residential co-ethnic density at the provincial level complicates the interpretation of the mechanism behind the effect that co-ethnic density may have on economic outcomes, analyzing the variation in co-ethnic population size across a broader geographical region helps alleviate obvious endogeneity concerns that may arise from what is known as the residential sorting problem (Duncan, 1957). Settlement patterns of migrants are hardly ever determined randomly, and it is likely that one engages with certain migratory networks or moves to a specific location for different unobserved reasons. For example, those who lack legal working rights or human capital in the form of education or language skills may opt for the protection of an ethnic enclave where they expect these factors to matter

less for their economic well-being. Pre-established migration links and family reunification also lead to the filtering of migrants from specific backgrounds to certain destinations. Thus, the clustering of similar types of people in the same spaces may bias estimates of the effect of co-ethnic density on earnings in function of the unobserved “quality” of the immigrant group being referenced (Schüller et al., 2022) (Kanas et al., 2012). I assume, however, that the residential sorting effect is essentially diluted as the level of covariance between the distribution of unobserved cohort “quality” and the distribution of co-ethnic density is reduced when measuring co-ethnic density at an aggregated geographic level as opposed to at a highly localized level.²¹ In fact, this is one argument employed in Cutler & Glaeser (1997) famous paper, *Are Ghettos Good or Bad?*, for why the authors decide to measure the impact of racial segregation across American cities rather than within them.

Indeed general differences in motives for migration are already taken into account by restricting the study population to the subset of economic immigrants in the ENI, but this does not quell all of the endogeneity concerns that arise from initial settlement decisions and uneven cohort effects. The migratory network itself becomes an actor in the integration and settlement of new cohorts of immigrants, sorting and distributing groups who choose to engage with the network for various different underlying reasons into different locations. To illustrate this and how it might be affecting the results seen in Table 4 independent of the labor stock argument already discussed at the beginning of this section, I return to the concept of chain migration. To oversimplify what merits an entire discussion that is out of the scope of this paper, for sake of argument I distinguish between the types of individuals that participate in these migratory chains as “leaders” and “followers”, the former representing the early pioneers that first set sail for Spain, or at least the first that found success in Spain, who, via social and economic networks, opened the doors for many family members, friends, acquaintances, community members, and fellow countrymen to follow. The selectivity that arises from this distinction is apparent and well documented in the literature.²² Pioneers tend to be younger, single men more able to get up and leave and find employment than, say, their female counterparts who have children and face greater restrictions in the labor market. Pioneers may also have more resources to make the trip over and find housing as migratory networks are not yet in place to assist with that, or may generally be more adaptable as the option of finding emotional comfort in an ethnic enclave does not exist yet. If this is the case, it suggests that early or small migration cohorts who may represent this positive selection bias have more tendency to concentrate than those who arrive as part of the chain that are more numerous and diffuse.

In addition, motivational differences between year-of-arrival cohorts may also have resulted in a type of measurement bias in the data. Up until the late 1980s, Spain had no sort of law

²¹Measuring at the province-level also prevents the calculation of outliers that may happen when measuring co-ethnic density in sparsely populated municipalities.

²²See Faist (2000), Lindstrom & López Ramírez (2010), and Wessendorf (2018) for some examples.

regulating immigration as it had never been a major destination for international migrants previously (Hooper, 2019). Expired contracts, bureaucracy, and growing unmatched demand for visa allotments in the construction, agricultural, and hospitality sectors drove up the number of irregular immigrants and in 2004 the regularization mechanism of *arraigo social* was established, granting legal status to foreigners who have been living in Spain for at least two years as long as they can prove that they have worked six months illegally in the country, or that they have Spanish family ties or have socially integrated in a sufficient manner (Gálvez Iniesta, 2020).²³ The incentives that this provides for registering in the local municipal system regardless of legal status (access to free medical care and public education is also conditional on being registered in the local municipal system), while improving authorities' and researchers' capacity to paint a more accurate picture of the immigrant situation in Spain, could also represent a change in underlying characteristics between those observed in the *Padrón* data prior to 2004 and those observed after. In this sense, a Cabo Verdean that settled in a province prior to 2004 with relatively higher density of Cabo Verdeans may indicate that the high relative co-ethnic density is actually just measuring areas where there are more established Cabo Verdeans who, say, may have legal residency or work permits and thus were not deterred from registering in the municipal system. Therefore, there may be other provinces with high relative densities of Cabo Verdeans, but these are not observed in the data prior to 2004 if, say, these groups were made up of disproportionately more illegal immigrants and therefore more disadvantaged.²⁴ Since return migration is of course an event that takes place with greater probability over time for certain immigrants over others, this too presents a possible selection bias in the data. However, very little can be done to address this issue.

Under this hypothesis, it could be the case that the measure of the effect of relative co-ethnic density on immigrant wages is biased upward if there is a positive relationship between the log of relative co-ethnic density and whether one settled in the province before 2004. However, when testing empirically this relationship, there is insufficient evidence to say that there is any significant effect of having arrived pre-2004 on the log of relative co-ethnic density ($t = 1.37$, $p = 0.17$). There is, however, a positive correlation between the logs of absolute and relative co-ethnic density ($r = 0.54$, $95\% \text{ CI} = [0.52, 0.55]$), and if I want to make any claims about the combined effect, $\beta_1 + \beta_2$, on the wages of someone who, say, has a small network as represented with the absolute measure, but a strong network as represented with the relative measure, then measurement bias could be a cause of concern since we know already that absolute co-ethnic density is positively correlated with year of arrival.

²³Achieving *arraigo* through the latter route requires proof of a work-contract offer that should last at least a year.

²⁴See Farelo & González (2008) for an intriguing and complete account of Cape Verdean migration in Spain.

4.2.2 IV Regression

To summarize all of the above, endogeneity in my study as specified thus far is inevitable. Some studies have exploited the pseudo-random variation in co-ethnic concentration due to dispersion policies that have been implemented for migrant refugees (Damm, 2009) (Tumen, 2015). However, in the case of unforced migration such quasi-experimental design does not exist and researchers must find an exogenous source of variation to use if they wish to identify the causal effect of co-ethnic density on immigrant wages. In the face of the non-randomness of networks and settlement patterns, I will instrument my co-ethnic density measures with their corresponding measures calculated from national census data from 1991. Just like the *Padrón* data, the census data provides total population counts by province as well as within-province counts stratified by country of origin from which densities are derived.

Two conditions must be met in order to achieve consistent estimates via instrumental variable regression. The first is that the instrument must be correlated with the endogenous explanatory variable at hand. With the rapid modernization of the economy from the end of the Franco regime into the 1980s came the expansion of urban industrial areas throughout the country and a growing concentration of people in Madrid, Barcelona, and other major metropolitan regions situated around the perimeters of the country (Gormsen & Klein, 1986). Naturally, many immigrants at the time also gravitated to these regions and continued to do so during the economic boom from 1998 to 2007. A comprehensive multivariate cluster analysis by Parreño-Castellano, Moreno-Medina, Domínguez-Mujica, & Santana Rivero (2021) shows that most immigrants, regardless of country-of-origin, gravitate towards the bigger urban municipalities in Spain. Therefore, I expect at least some correlation between the co-ethnic density measures from the *Padrón* data and the 1991 census data, driven mainly by the fact that immigrants in general will tend to settle in cities where jobs are, regardless of any pre-existing migratory chains that may have led them there. First stage regression modelling in Models 2 and 3 of Table 6 shows that there is indeed strong correlations between the density measures in 1991 and their respective measures in the ENI population a decade later.

The second condition that must be fulfilled is that the instrument, conditional on the other exogenous variables in the model, is not correlated with the error term. While there is a plethora of endogeneity issues that may be putting the instrument at risk of the same concerns, the most immediate issue is that of selection bias. As discussed in the previous section, it is likely that immigrants self-select into networks of varying quality, and these networks then distribute them to certain regions over others where they experience certain economic outcomes not based on where they settle, but rather based on differences in motives, preferences, or ability that selected them into those networks in the first place. In 1991, however, there were very few foreigners living in Spain, let alone any chain migratory

networks that had been established at that point. For much of the 20th century, Spain's migration outflows outpaced its inflows. Prior to 1974, many Spaniards were leaving the country to work abroad under the guest-worker program that was implemented to fill the labor shortages during the post-war boom in Western Europe (Castles, 1986). Emigration rates dropped during the next decade while foreign inflows kept a slow but steady pace, maintaining an almost net-zero migration rate between the years 1976 and 1988. By 1990, still only 2% of the Spanish population was made up of foreign nationals, many of whom were spending a temporary amount of time in Spain as they transitioned to more popular immigrant destinations such as France or Germany (Pérez, 2003). Immigrant inflow to Spain at this point was clearly not anywhere close to the scale that it would see in the 2000s.

However, the intersection of a global economic crisis, closing of borders of long-time immigrant importing countries, strategic geo-economic positioning, and recent admittance into the European Union turned Spain from a “waiting room” into a new migrant destination at the start of the 1990s. With the signing of the Treaty of Accession came also the enactment of the country's first national immigration law, the *Ley de extranjería* of 1986. However, the policy lacked teeth as it viewed immigration as a temporary issue, something that needed to be done to deal with those foreigners already living in Spain in the face of the newly minted EU membership card. The 1996 amendment of the national immigration law signaled the recognition by the Spanish government of immigration as a structural phenomenon. Liberalized regularization procedures, extension of immigrant rights, and implementation of work permit quota systems organized both internal and external migration flows in the late 1990s into a political-economic system to which migratory networks responded to. Thus, I argue that the 1991 census came at a critical point in Spanish immigration history. There were enough foreigners living in Spain at the time that counts by province and country of origin group were non-zero allowing temporal correlations to be drawn between the co-ethnic density measures, but the counts were very few as migration chains were hardly established at that point. Any network that would have been established would have been small, and the temporal distance between them and networks of immigrants arriving in the 2000s ensures independence of our instrument from the endogenous explanatory variable.²⁵

²⁵We see in Figure 4 of the Appendix, less of those immigrants who arrived to Spain in 1990 had been influenced by co-ethnic contacts who had previous emigrated to Spain. It isn't until 1997 that we see for the first time a majority of the yearly arrival cohort having migrated under network influence, this influence increasing substantially in later cohorts.

Table 6: Effect of log of co-ethnic density (absolute and relative) on log of hourly wage, OLS and 2SLS

| | Log of hourly wage OLS | logCOED (absolute) 2SLS - 1st stage | logCOED (relative) | Log of hourly wage 2SLS - 2nd stage |
|-------------------------|---------------------------|--|----------------------|--|
| | (1) | (2) | (3) | (4) |
| logCOED (absolute) | −0.020*** (0.006) | | | −0.021 (0.032) |
| logCOED (relative) | 0.015 (0.009) | | | 0.036 (0.031) |
| logCOED1991 (absolute) | | 0.906*** (0.162) | −0.185** (0.093) | |
| logCOED1991 (relative) | | −0.372** (0.163) | 0.724*** (0.094) | |
| Female | −0.206*** (0.028) | −0.027 (0.030) | −0.049*** (0.017) | −0.220*** (0.029) |
| Years in Spain | 0.012*** (0.002) | −0.092*** (0.004) | −0.005** (0.002) | 0.011*** (0.003) |
| Secondary (1st cycle) | −0.006 (0.018) | −0.051 (0.047) | −0.042 (0.027) | −0.009 (0.018) |
| Secondary (2nd cycle) | 0.022 (0.014) | −0.072* (0.038) | −0.039* (0.022) | 0.023 (0.014) |
| Tertiary (1st cycle) | 0.102*** (0.019) | −0.092* (0.051) | −0.086*** (0.029) | 0.098*** (0.019) |
| Tertiary (2nd cycle) | 0.198*** (0.057) | −0.312* (0.159) | −0.216** (0.092) | 0.219*** (0.058) |
| Age | 0.014*** (0.004) | −0.059*** (0.008) | −0.003 (0.005) | 0.015*** (0.004) |
| Age squared | −0.0002*** (0.0001) | 0.001*** (0.0001) | 0.00003 (0.0001) | −0.0002*** (0.0001) |
| Sector controls | Yes | Yes | Yes | Yes |
| Origin country controls | Yes | Yes | Yes | Yes |
| Region controls | Yes | Yes | Yes | Yes |
| Observations | 6,525 | 6,525 | 6,525 | 6,525 |
| R ² | | 0.527 | 0.511 | |
| Adjusted R ² | | 0.515 | 0.499 | |
| Residual Std. Error | | 1.155 | 0.664 | |
| F Statistic | | 44.090*** | 41.333*** | |

Note:

*p<0.1; **p<0.05; ***p<0.01

Results shown for wage estimations in (1) and (4) are based on the Heckman selection model.

Table 6 presents the results of the 2SLS regression that instruments both co-ethnic density measures in province p in the year in which the individual of reference began living there with the respective measures from the same province in 1991 (Model 1).²⁶ The effects, which can be understood as consistent estimates of the parameters of interest so long as the assumption holds that the instruments only act on the response variable through the endogenous variables, show a similar effect for absolute co-ethnic density as the one estimated by OLS (Model 4). For relative co-ethnic density, we see an even greater positive effect, though both

²⁶Some countries in 1991 did not map perfectly to countries in 2007 due to the dissolution of the Soviet Union and various other political happenings. Country mappings can be seen in Table 9 of the 6.

estimates are statistically insignificant at $\alpha = 0.1$. Still, the results are convincing and provide evidence that network strength, regardless of the other covariates, has positive effects on immigrant wages while network magnitude has the opposite.

5 CONCLUSION

This analysis has attempted to identify and estimate the independent effect of residential settlement patterns on the wages of economic immigrants in Spain. However, like any other social phenomenon, there are meso-level factors that need to be considered, but separating the individual from the whole can be a daunting task. In this case of my study, co-ethnic density has come to represent immigrant networks since the non-random decision to settle in one region over another is often driven by where other similar individuals have settled before. Though unforced migratory movements present a great issue of selection bias into certain networks, assessing critically the endogeneity concerns has actually proven to be a rather useful exercise in understanding the different dimensions of migratory networks. By controlling for both absolute and relative co-ethnic densities, I disentangle the negative effects of increased immigrant labor stock and the positive effects of actively leveraging migratory networks on immigrant wages. My findings point to the conclusion that networks are beneficial when they are strong, but not necessarily when they are big.

6 Appendix

Table 7: Comparison of Distributions of Age and Tenure in Spain for Economic and Non-Economic Immigrants

| Variable | Economic Immigrants | | | Non-economic Immigrants | | | Test of Equal Means |
|----------------|---------------------|-------|-------|-------------------------|-------|-------|---------------------|
| | Mean | SD | Count | Mean | SD | Count | p-value |
| Years in Spain | 7.33 | 8.8 | 8572 | 18.34 | 17.7 | 6815 | 0 |
| Age | 36.62 | 11.17 | 8569 | 42.36 | 16.91 | 6855 | 4.4e-126 |

Table 8: Comparison of Cohorts of Wage Earning and Non-Wage Earning Economic Immigrants

| Variable | Wage Earners | | | Non-Wage Earners | | | Test of Equal Props. |
|---|--------------|-------|---------|------------------|-------|---------|----------------------|
| | Count | Total | Percent | Count | Total | Percent | p-value |
| Gender | | | | | | | |
| Male | 2230 | 4135 | 53.93 | 1032 | 2390 | 43.18 | 7.33e-17 |
| Female | 1905 | 4135 | 46.07 | 1358 | 2390 | 56.82 | 7.33e-17 |
| Level of studies | | | | | | | |
| No secondary | 1123 | 4135 | 27.16 | 694 | 2390 | 29.04 | 0.109 |
| Secondary (1st cycle) | 627 | 4135 | 15.16 | 423 | 2390 | 17.7 | 0.00804 |
| Secondary (2nd cycle) | 1757 | 4135 | 42.49 | 903 | 2390 | 37.78 | 0.000213 |
| Tertiary (1st cycle) | 588 | 4135 | 14.22 | 352 | 2390 | 14.73 | 0.599 |
| Tertiary (2nd cycle) | 40 | 4135 | 0.97 | 18 | 2390 | 0.75 | 0.452 |
| Country classification | | | | | | | |
| Andean countries | 1495 | 4135 | 36.15 | 711 | 2390 | 29.75 | 1.58e-07 |
| Asia and Oceania | 101 | 4135 | 2.44 | 61 | 2390 | 2.55 | 0.848 |
| Developed countries (>75% GDPpc of Spain) | 148 | 4135 | 3.58 | 164 | 2390 | 6.86 | 3.08e-09 |
| Eastern Europe, Balkans, other European nationalities | 1096 | 4135 | 26.51 | 569 | 2390 | 23.81 | 0.0174 |
| Northern Africa & Middle East | 443 | 4135 | 10.71 | 343 | 2390 | 14.35 | 1.63e-05 |
| Other Latin American countries | 734 | 4135 | 17.75 | 447 | 2390 | 18.7 | 0.353 |
| Sub-Saharan Africa (excl. S. Africa) | 118 | 4135 | 2.85 | 95 | 2390 | 3.97 | 0.0172 |
| Married woman with kids? | | | | | | | |
| No | 3523 | 4135 | 85.2 | 1812 | 2390 | 75.82 | 4.35e-21 |
| Yes | 612 | 4135 | 14.8 | 578 | 2390 | 24.18 | 4.35e-21 |

Table 9: Countries Represented in the ENI and Mappings to Countries in Padrón data and 1991 Census

| LITERAL | PADRÓN | 1991.CENSUS | GROUPING |
|----------------------|----------------------|----------------------|---|
| Bulgaria | Bulgaria | Bulgaria | Eastern Europe, Balkans, other European nationalities |
| Francia | Francia | Francia | Developed countries (>75% GDPpc of Spain) |
| Italia | Italia | Italia | Developed countries (>75% GDPpc of Spain) |
| Polonia | Polonia | Polonia | Eastern Europe, Balkans, other European nationalities |
| Portugal | Portugal | Portugal | Developed countries (>75% GDPpc of Spain) |
| Reino Unido | Reino.Unido | Reino.Unido | Developed countries (>75% GDPpc of Spain) |
| Alemania | Alemania | Alemania | Developed countries (>75% GDPpc of Spain) |
| Rumanía | Rumanía | Rumanía | Eastern Europe, Balkans, other European nationalities |
| Ucrania | Ucrania | U.R.S.S. | Eastern Europe, Balkans, other European nationalities |
| Moldova | Moldova | U.R.S.S. | Eastern Europe, Balkans, other European nationalities |
| Georgia | Georgia | U.R.S.S. | Eastern Europe, Balkans, other European nationalities |
| Lituania | Lituania | U.R.S.S. | Eastern Europe, Balkans, other European nationalities |
| Armenia | Armenia | U.R.S.S. | Eastern Europe, Balkans, other European nationalities |
| Rusia | Rusia | U.R.S.S. | Eastern Europe, Balkans, other European nationalities |
| Argelia | Argelia | Argelia | Northern Africa & Middle East |
| Gambia | Gambia | Gambia | Sub-Saharan Africa (excl. S. Africa) |
| Ghana | Ghana | Ghana | Sub-Saharan Africa (excl. S. Africa) |
| Guinea | Guinea | Guinea | Sub-Saharan Africa (excl. S. Africa) |
| Guinea Ecuatorial | Guinea.Ecuatorial | Guinea.Ecuatorial | Sub-Saharan Africa (excl. S. Africa) |
| Mali | Mali | Mali | Sub-Saharan Africa (excl. S. Africa) |
| Marruecos | Marruecos | Marruecos | Northern Africa & Middle East |
| Mauritania | Mauritania | Mauritania | Sub-Saharan Africa (excl. S. Africa) |
| Nigeria | Nigeria | Nigeria | Sub-Saharan Africa (excl. S. Africa) |
| Senegal | Senegal | Senegal | Sub-Saharan Africa (excl. S. Africa) |
| México | México | México | Other Latin American countries |
| Cuba | Cuba | Cuba | Other Latin American countries |
| Honduras | Honduras | Honduras | Other Latin American countries |
| Nicaragua | Nicaragua | Nicaragua | Other Latin American countries |
| República Dominicana | República.Dominicana | República.Dominicana | Other Latin American countries |
| Argentina | Argentina | Argentina | Other Latin American countries |
| Bolivia | Bolivia | Bolivia | Andean countries |
| Brasil | Brasil | Brasil | Other Latin American countries |
| Colombia | Colombia | Colombia | Andean countries |
| Chile | Chile | Chile | Other Latin American countries |
| Ecuador | Ecuador | Ecuador | Andean countries |
| Paraguay | Paraguay | Paraguay | Other Latin American countries |
| Perú | Perú | Perú | Andean countries |
| Uruguay | Uruguay | Uruguay | Other Latin American countries |
| Venezuela | Venezuela | Venezuela | Other Latin American countries |
| China | China | China | Asia and Oceania |
| Filipinas | Filipinas | Filipinas | Asia and Oceania |
| India | India | India | Asia and Oceania |
| Pakistán | Pakistán | Pakistán | Asia and Oceania |

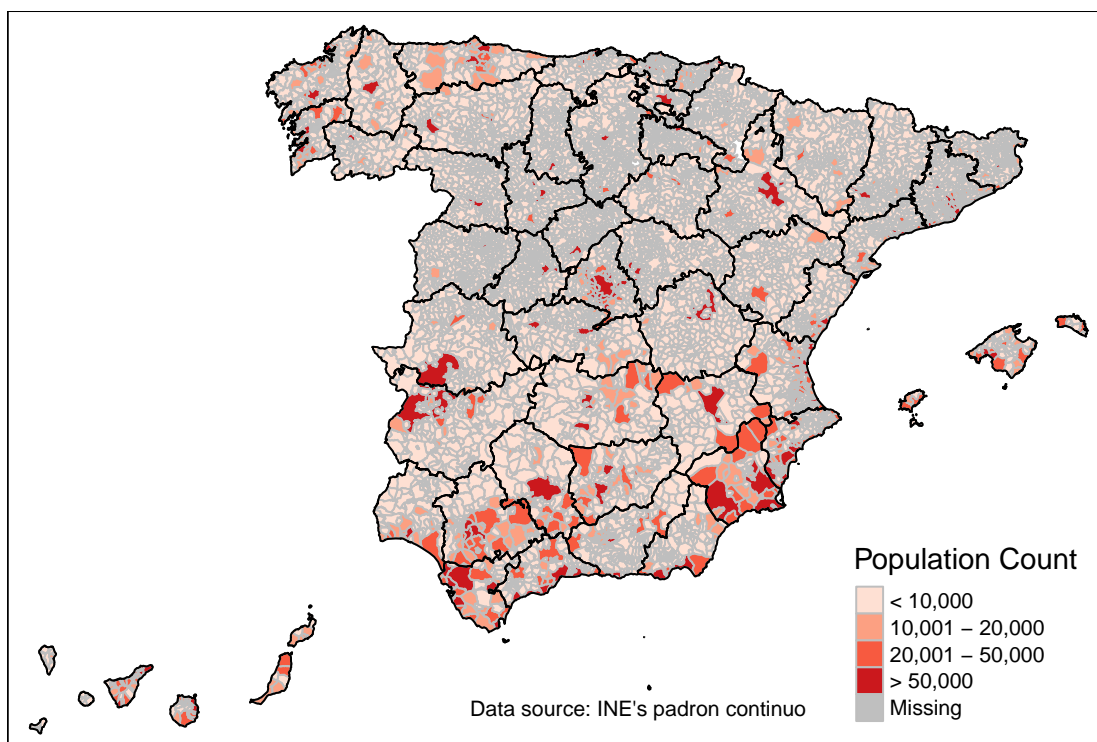


Figure 3: Municipalities by Population Count (2007)

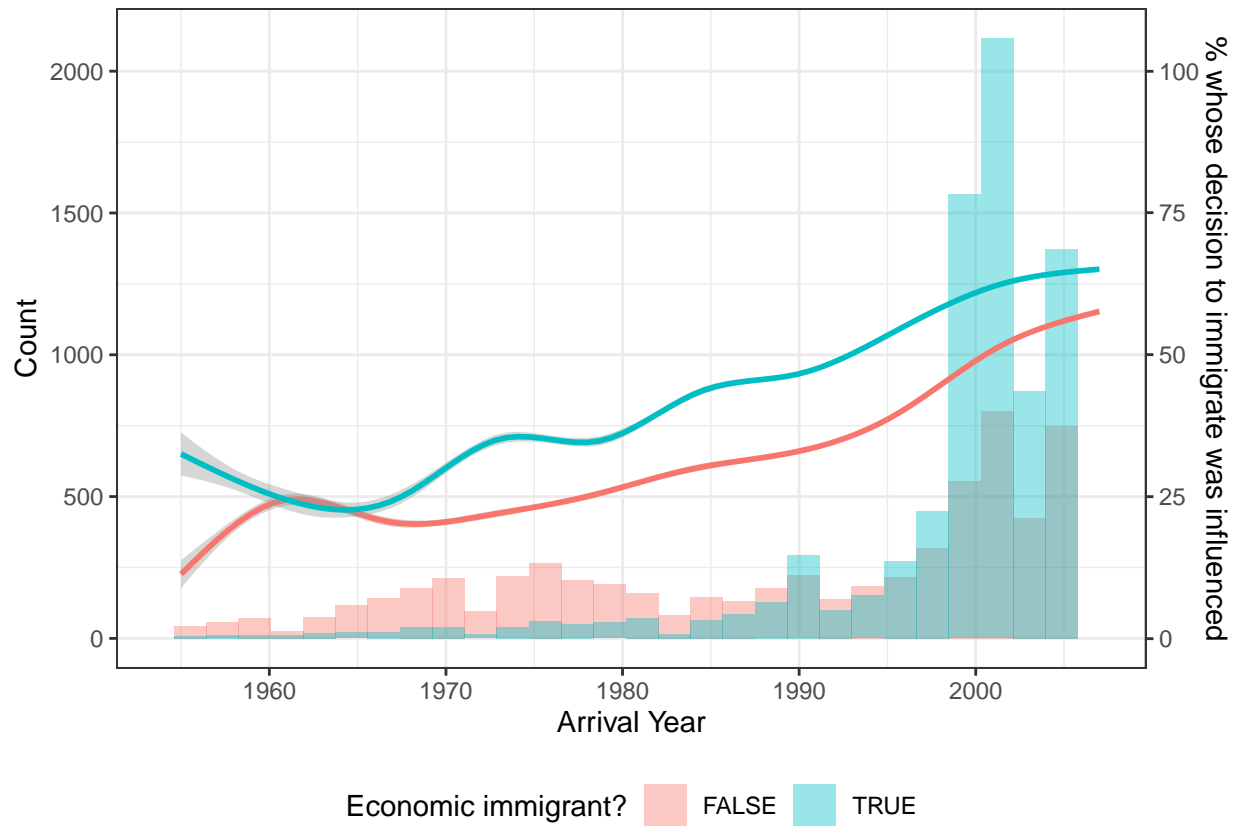


Figure 4: Counts and percentage of ENI respondents whose decision to migrate to Spain was influenced by a family member, friend, or neighbor from home country who had already emigrated to Spain, by arrival year (1954 - 2007)

7 References

- Borjas, G. J. (1994). Ethnicity, neighborhoods, and human capital externalities. National Bureau of Economic Research Cambridge, Mass., USA.
- Borjas, G. J. (2000). Ethnic enclaves and assimilation. *Swedish Economic policy review*, 7(2), 89–122. ECONOMIC COUNCIL OF SWEDEN.
- Borjas, G. J., Freeman, R. B., Katz, L. F., DiNardo, J., & Abowd, J. M. (1997). How much do immigration and trade affect labor market outcomes? *Brookings papers on economic activity*, 1997(1), 1–90. JSTOR.
- Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3, 1801–1863. Elsevier.
- Castles, S. (1986). The guest-worker in western europe—an obituary. *International migration review*, 20(4), 761–778. SAGE Publications Sage CA: Los Angeles, CA.
- Chiswick, B. R. (1978). The effect of americanization on the earnings of foreign-born men. *Journal of Political Economy*, 86(5), 897–921. JSTOR.
- Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2001). The determinants of the geographic concentration among immigrants: Application to australia. *Australasian Journal of Regional Studies*, The, 7(2), 125–150.
- Chiswick, B. R., & Miller, P. W. (2005). Do enclaves matter in immigrant adjustment? *City & Community*, 4(1), 5–35. SAGE Publications Sage CA: Los Angeles, CA.
- Cutler, D. M., & Glaeser, E. L. (1997). Are ghettos good or bad? *The Quarterly Journal of Economics*, 112(3), 827–872. MIT Press.
- Damm, A. P. (2009). Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence. *Journal of Labor Economics*, 27(2), 281–314. The University of Chicago Press.
- Doeringer, P. B., & Piore, M. J. (1975). Unemployment and the dual labor market. *The Public Interest*, 38, 67. National Affairs, inc.
- Dorigo, G., & Tobler, W. (1983). Push-pull migration laws. *Annals of the Association of American Geographers*, 73(1), 1–17. Taylor & Francis.
- Duncan, O. D. (1957). The measurement of population distribution. *Population Studies*, 11(1), 27–45. Taylor & Francis.
- Epstein, G. S., & Gang, I. N. (2010). Migration and culture. *Migration and culture*. Emerald Group Publishing Limited.
- Faist, T. (2000). 143The Selective and Diffusion Functions of Social Capital: Why So Many Migrants Out of So Few Places? *The Volume and Dynamics of International Migration and Transnational Social Spaces*. Oxford University Press. Retrieved from <https://doi.org/10.1093/acprof:oso/9780198293910.003.0006>
- Farelo, R. M., & González, L. O. (2008). Cape verdeans in spain. *Transnational Archipelago*, 73.

- Friedberg, R. M. (2000). You can't take it with you? Immigrant assimilation and the portability of human capital. *Journal of labor economics*, 18(2), 221–251. The University of Chicago Press.
- Gálvez Iniesta, I. (2020). The size, socio-economic composition and fiscal implications of the irregular immigration in Spain.
- Gormsen, E., & Klein, R. (1986). Recent trends of urban development and town planning in Spain. *GeoJournal*, 13(1), 47–57. Springer.
- Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *The American economic review*, 60(1), 126–142. JSTOR.
- Hasmath, R. (2012). The ethnic penalty: Immigration, education and the labour market. Burlington, VT and Surrey. Ashgate, UK.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161. JSTOR.
- Hooper, K. (2019). Spain's labour migration policies in the aftermath of economic crisis. *Report April, Migration Policy Institute, Brussels*.
- Hou, F., Lu, Y., & Schimmele, C. (2021). Trends in immigrant overeducation: The role of supply and demand. *International Migration*, 59(3), 192–212. Wiley Online Library.
- INE (Ed.). (2021). Population figures at 1 July 2021 migration statistics. First half of 2021.
- INE (Ed.). (2022). La población inscrita en España en el padrón continuo aumenta en 50.490 personas durante 2021 y se sitúa en 47,4 millones.
- Joona, P. A., Gupta, N. D., & Wadensjö, E. (2014). Overeducation among immigrants in Sweden: Incidence, wage effects and state dependence. *IZA Journal of Migration*, 3(1), 1–23. Springer.
- Kanas, A., Chiswick, B. R., Van Der Lippe, T., & Van Tubergen, F. (2012). Social contacts and the economic performance of immigrants: A panel study of immigrants in Germany. *International Migration Review*, 46(3), 680–709. SAGE Publications Sage CA: Los Angeles, CA.
- Kanas, A., & Van Tubergen, F. (2009). The impact of origin and host country schooling on the economic performance of immigrants. *Social Forces*, 88(2), 893–915. The University of North Carolina Press.
- Kerr, W. R., & Mandorff, M. (2015). *Social networks, ethnicity, and entrepreneurship*. National Bureau of Economic Research.
- Kuznets, S. (1960). Economic structure and life of the Jews. In L. Finkelstein (Ed.), *The minority members: History, culture, and religion* (pp. 14–32). Publication Society of America.
- Le-Quang, H., & Vallizadeh, E. (2022). The returns to school-quality-adjusted education of immigrants in Germany. *Journal for Labour Market Research*, 56(8).
- Light, I., Kwun, I. J., & Zhong, D. (1990). Korean rotating credit associations in Los Angeles. *Amerasia Journal*, 16(2), 35–54. Taylor & Francis.

- Lindstrom, D. P., & López Ramírez, A. (2010). Pioneers and followers: Migrant selectivity and the development of US migration streams in latin america. *The Annals of the American Academy of Political and Social Science*, 630(1), 53–77. Sage Publications Sage CA: Los Angeles, CA.
- Liu, L. Q. (2018). *Regional labor mobility in spain*. International Monetary Fund.
- MacDonald, J. S., & MacDonald, L. D. (1964). Chain migration ethnic neighborhood formation and social networks. *The Milbank Memorial Fund Quarterly*, 42(1), 82–97. JSTOR.
- Massey, D. S., & España, F. G. (1987). The social process of international migration. *Science*, 237(4816), 733–738. American Association for the Advancement of Science.
- Maza, A., & Villaverde, J. (2009). Provincial wages in spain: Convergence and flexibility. *Urban Studies*, 46(9), 1969–1993. SAGE Publications Sage UK: London, England.
- McAuliffe, M., & Khadria, B. (2020). World migration report 2018. International Organization for Migration.
- Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics*, 118(2), 549–599. MIT Press.
- Ozgen, C., Nijkamp, P., & Poot, J. (2012). Immigration and innovation in european regions. *Migration impact assessment*. Edward Elgar Publishing.
- Parreño-Castellano, J. M., Moreno-Medina, C., Domínguez-Mujica, J., & Santana Rivero, C. (2021). Mapping foreign immigration in spain (1998–2018). Trends and spatial patterns. *Journal of Maps*, 17(1), 79–84. Taylor & Francis.
- Pérez, N. O. (2003). Spain: Forging an immigration policy. *Migration Information Source*.
- Portes, A. (1981). 13 modes of structural incorporation and present theories of labor immigration. *International Migration Review*, 15(1_suppl), 279–297. SAGE Publications Sage CA: Los Angeles, CA.
- Portes, A., & Bach, R. L. (1985). *Latin journey: Cuban and mexican immigrants in the united states*. Univ of California Press.
- Portes, A., & Shafer, S. (2007). Revisiting the enclave hypothesis: Miami twenty-five years later. *The sociology of entrepreneurship*. Emerald Group Publishing Limited.
- Rabe, B., & Taylor, M. P. (2012). Differences in opportunities? Wage, employment and house-price effects on migration. *Oxford Bulletin of Economics and Statistics*, 74(6), 831–855. Wiley Online Library.
- Reher, D., Cortés, L., González, F., Requena, M., Sánchez, M. I., Sanz, A., & Stanek, M. (2008). Informe encuesta nacional de inmigrantes (ENI-2007). INE, Documentos de Trabajo.
- Reher, D., & Requena, M. (2009). The national immigrant survey of spain: A new data source for migration studies in europe. *Demographic Research*, 20, 253–278. JSTOR.
- Sanders, J. M., & Nee, V. (1987). Limits of ethnic solidarity in the enclave economy. *American sociological review*, 745–773. JSTOR.
- Schüller, S. C. et al. (2022). Ethnic enclaves and immigrant economic integration. *IZA*

World of Labor.

Tumen, S. (2015). The use of natural experiments in migration research. *IZA World of Labor*.

Wessendorf, S. (2018). Pathways of settlement among pioneer migrants in super-diverse london. *Journal of Ethnic and Migration Studies*, 44(2), 270–286. Taylor & Francis.