ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Specialisation International Economics

Spanish Emigration

A Panel Data Analysis (2002-2013)

Author: Lorena Gomez Cruz
Student number: 435222
Thesis supervisor: Prof. dr. J.M.A. Viaene
Year 2015/2016

Abstract: international migration is a topic of vital importance, as Spanish population abroad has increased 56% from 2002 to 2013, with currently circa 2.3 million Spanish citizens living abroad. Since this crisis hit Spain in 2008, GDP pc growth plummeted, inequality increased and unemployment reached 26.9% in 2013, with some sources registering up to 400,000 emigrants per year. Although non-nationals emigration outweighs nationals' emigration, the second has caught more attention, generating alarm of a possible "brain drain" and "lost generation". I study this phenomenon using a gravity model that contains data of emigration flows from Spain to 60 destination countries from 2002 to 2013, finding that relative GDP per capita, distance, population, relative unemployment and relative inequality all play a significant role in these flows. However, welfare variables and others referring to relations between countries are mainly not significant.

Keywords: *international migration, gravity model, Spain, crisis.*

Table of Contents

List of Tables and Graphs

1. Introduction

Nowadays international migration is a topic of vital importance, as around 3% of the world population lives in a country other than the one they were born in. Although it may sometimes be discredited as a low number (after all, it is "only" around 232 million people out of over 7 billion), it has increased over 50% in the last 50 years (Felbermayr et al., 2014). In the case of Spain, it is an even higher rise: 56%[1] from 2002 to 2013, with currently circa 2.3 million Spanish citizens living abroad[2].

During the last century, Spain's migration history has changed dramatically, from high emigration rates to net entries of immigrants, and in the last few years, back to high emigration rates and with net migration in negative values since 2011 (Bertoli et al., 2013). As many of these changes have taken place since the beginning of this century, this paper analyses the causes of emigration in Spain from 2002 to 2013 to a set of 60 European, Latin American and North African countries, focusing mainly on income inequality, unemployment and welfare.

Throughout the XXth century Spain was mainly exporting labour, and numbers were especially high in the first decades. Spanish emigrants chose mainly South American countries, and their reasons to move out of their country were varied: the rapid growth of population in Spain, economic stagnation or an unstable political situation. However, due to wars and the subsequent establishment of a dictatorship, emigration was banned and rates declined sharply until the 1960s, when reforms took place in order to open Spain again to the world. This boom in emigration (mainly to countries with shortage of unskilled workers like France, Switzerland, Germany and Venezuela), which saw rates of up to 7 per 1000 of total population, declined again as economic development consolidated (Izquierdo et al., 2015).

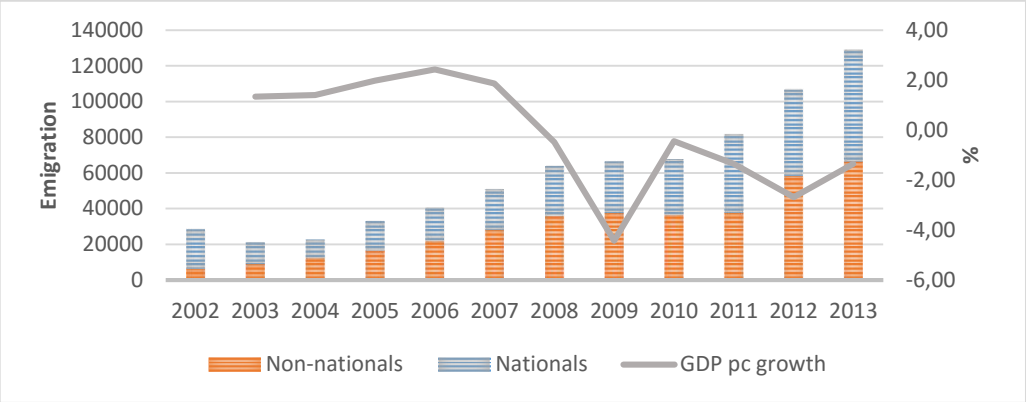
In fact, after Spain joined the European Economic Community in 1986, net migration was positive, meaning that the country received large numbers of immigrants. As Izquierdo et al. (2015) show, in each of the years previous to the Great Recession (and even the first years of the crisis), Spain received an average of 1.4% of its total population until they reached 14% of it[3], while outflows were much smaller. Moreno Torres and

[1] According to Consulates data (CERA).
[2] According to Padrón de Españoles Residentes en el Extranjero.
[3] OECD: World Migration in Figures 2013

López Casasnovas (2006) note that immigrants started to arrive from Africa (mostly from Morocco) at the end of the 1980s, in contrast with the traditional European and South American (mostly Peruvian and Bolivian) immigration. These immigrants influenced the composition of the labour force, as well as geographical mobility, and they are also important when attempting to explain a great proportion of the current emigration rates.

Since the crisis hit Spain in 2008, however, European and American immigration has declined significantly. The situation in Spain worsened as GDP pc growth plummeted in 2009, inequality increased and unemployment reached a peak of 26.9% (55% for young people) in 2013 (see Graphs 1.1 and 1.2). As immigration fell, emigration started to increase. Although data for migration outflows is not consistent across sources (see more in Section 3.1.1), with some of them registering up to 400,000 emigrants per year, they all agree on the magnitude of the phenomenon and the shift in migration trends. It is vital to notice, though, that a great portion of this emigration is composed of non-nationals, that is, mainly former immigrants that return to their countries. Although non-nationals emigration outweighs nationals' emigration, the second has caught more attention due to its novelty (Domingo i Valls et al., 2014). Since 2012 outflows offset inflows, which had not happened in Spain since the 1970s, not only for non-nationals, but also for nationals, who mostly emigrate to Europe and the USA (Izquierdo et al., 2015). As Ramos and Suriñach (2013) point out, the new emigration has generated alarm and worry in the country as many of the migrants are young highly qualified people (who are called by some the "lost generation"), which could have negative effects for the economy and society.

Graph 1.1 – GDP pc growth and emigration[4] in Spain (2002-2013)



Source: INE, Residential Variation Statistics and World Bank

Graph 1.2 – Inequality and unemployment in Spain (2004-2012)

Source: World Bank

These new Spanish emigration trends have not been researched broadly, due to them being so recent, although repercussion in the media has been wide. In fact, the only global study, carried out by the Spanish Government, has not been published[5] (Rodríguez-Fariñas et al., 2016). Nevertheless, there are some published studies on immigration, such as Moreno Torres and López Casasnovas (2006); immigration and emigration, such as Izquierdo et al. (2015), which is one of the most complete, although they only focus on education and unemployment for emigration between 2008 and 2012; Navarrete Moreno et al. (2014), which is based mainly on surveys and describes only emigration statistics for a few destination countries; and Rodríguez-Fariñas et al. (2016), which refers only to

[4] Only to the countries used in our empirical specification.
[5] Dirección General de Migraciones del Gobierno de España: *La movilidad exterior de los españoles y las españolas en la actualidad*.

Mexico and does not include an econometrics approach. None of them include inequality as an important determinant of emigration.

As for other broad studies on migration, there is plenty of literature that studies the factors that influence migration flows and inequality, either empirically of theoretically, such as Mayda (2010) or Borjas (1987), while others focus on the effect of welfare (Enache and Pânzaru, 2012). This will be expanded in Section 2.

With this paper I aim to expand the study by Izquierdo et al. (2015) by adding a longer timespan to the research, from 2002 to 2013, focusing on European and American countries. The time is limited due to the lack of data on emigration flows with a known destination country, which adds to the discrepancies in the recorded data. This will be addressed in Section 3.1. The approach, however, will be different to that of Izquierdo et al. (2015), as it will be based in the gravity model used by authors such as Lewer and Van den Berg (2008) and Mayda (2010). As inequality and welfare as variables that influence emigration have not been studied in depth for Spain, I will include them in this paper.

The results of the gravity model are not all fully conclusive. While the benchmark model, which is based on the effects of GDP per capita, distance and population, is always significant and has the correct signs (higher relative GDP per capita in Spain decreases emigration flows, same as distance, while bigger populations increase those flows). The same does not happen with all the variables of the augmented model, which are divided in three blocks: an augmented model that adds unemployment and relative deprivation, welfare and variables related to countries' relations. Higher relative unemployment in Spain increases emigration flows, while higher relative inequality decreases them. This last variable has a different sign than in the literature as Spain already has high inequality levels. However, both the welfare and relations blocks do not have the expected significant results, suggesting that there should be more research.

The paper will be organised as follows: in Section 2 I review the main findings regarding migration theory as well as the main empirical studies and the hypotheses that will be tested based on this review; Section 3 develops the methodology, including the empirical model and the data (including explanations of the most important variables for the Spanish case); in Section 4 I will explain the main results of the econometric specification, as well as the policy implications, and in Section 5 I will conclude.

2. Literature Review

In this section I will review the main existing literature on migration and its causes. From the main literature findings, I will extract the different hypotheses that will be tested using the gravity model described in Section 3.2. It is necessary to acknowledge the importance of two books in the migration literature: the *Handbook of Development Economics* and, especially, the *Handbook of the Economics of International Migration*. Both of them gather the most prominent papers published in their respective fields, and are the best sources of information regarding international migration.

This section will be divided in three sub-sections: first I will discuss the theoretical framework of this paper by reviewing the main ideas that have been developed in the field of international migration, especially the latest ones regarding the human capital approach and relative deprivation. Then I will summarize the main empirical findings of these theories, including the research that has been done for Spain's international migration. Finally, I will explain the hypotheses that can be obtained based in these first two sections.

2.1. Theoretical framework

The main migration models can be divided in two categories. In one, the models used to analyse the determinants of international migration, and in the other, the ones that determine its impact (Son and Noja, 2012). Other classifications include those by Greenwood (1985), who reviews different types of migration models, and states that economic models of migration rely on one of two perspectives: disequilibrium or equilibrium. The first one was very popular before the 1970s, and assumed that there were spatial differences that raised opportunities for utility gains that could be realized through migration. The other perspective assumes that the differences are compensated and there are no utility gains. As the main objective of this paper is to determine the causes of Spanish emigration, I will only focus on the models that deal with the causes of emigration, but including conclusions from both disequilibrium and equilibrium perspectives.

Bodvarsson et al. (2014) summarize the theories related to migration, although before the 1960s (Sjaastad, 1962) there was very little literature. However, it laid the ground for later analysis and formal development, starting with Adam Smith, who suggested that migration results from spatial disequilibrium in labour markets and returns to labour

supply. He also observed that, despite wage differences, migration flows are small: this is because costs are high compared to the gains. Already in the XXth century, Greenwood reviewed what are known as the "laws of migration": migrants usually move a short distance to large cities; these migrants are from rural areas, and their move increases gaps in rural population, causing those who live even further to migrate; out-migration and in-migration are inversely related; big migration waves generate compensating counter-waves and women are more likely to migrate than men. Furthermore, the early stages of the gravity model of migration were developed by Zipf (1946), who hypothesized that, while the volume of migration between two places is inversely proportional to distance, it is directly proportional to the product of the two populations.

The core of migration theory, and therefore the human capital approach, however, has always remained the same: migration takes place when individuals maximize their return to human capital investment by responding to spatial differences in labour market opportunities when those opportunities compensate the cost of relocating. Moreno Torres and López Casasnovas (2006) call this the neoclassical migration theory, which focuses on why migration takes place from a macroeconomic and microeconomic point of view. According to the first one, individuals respond to opportunities in countries taking into account migration costs, which can be material (transport) and psychological (leave the family behind). These opportunities are mainly real salary (nominal for the Keynesian theory) differences due to differences in supply and demand of labour: where labour is more abundant than capital, salaries will be lower, and the other way around. Therefore, workers will have an incentive to migrate to a region or country where salaries are higher until salaries are equalized (including migration costs). Some authors, though, defend that these differences cannot be explained only through costs, and others include as a cost the probability of being unemployed not only in the origin country, but also in the destination country. From a microeconomic point of view, individuals are rational and their decision to migrate results from a benefit-cost analysis, that is, they will migrate when their net returns are positive, which includes the unemployment rates.

The main conclusion that can obtained from this is that differences in GDP per capita (which is a proxy for wages or earnings) in origin (Spain in this case) and destination influence migration, that is, people will emigrate where GDP per capita is higher.

As Bodvarsson et al. (2014) write, Sjaastad (1962) was the first one to address the human capital approach to migration as we know it today by adding time. Although he did not provide a mathematical model, he asserted that migrants choose the destination that maximizes the present value of their lifetime earnings. The model, which would look like this, is adapted to the case that will be analysed in this thesis (emigration from Spain):

$$Net\ return = \sum_{t=1}^{T} \frac{(E_t^d - E_t^{Sp})}{(1+r)^t} - \sum_{t=1}^{T} \frac{(C_t^d - C_t^{Sp})}{(1+r)^t} \quad (2.1)$$

where "d" refers to the destination countries and "Sp" to the origin country, Spain. "E" measures the earnings in each period in the destination and origin countries, while "C" measures the costs of moving, which are assumed to be proportional to the distance between Spain and the destination. All of this while taking into account the rate of time preference (discount rate "r"). If the net return (which is the present value) is positive, the individual will migrate.

It is necessary to acknowledge that here we can observe two factors known as "pull" (opportunities in the destination) and "push" (deteriorating opportunities at home), that Sjaastad (1962) assumes are symmetrical, although the empirical literature has found the first type to be dominant. These "pull" and "push" factors are also part of the dual labour market theory (Moreno Torres and López Casasnovas, 2006), which states that international migration results from labour demand in developed countries (pull factors) instead of low wages or unemployment in poorer countries (push factor). This is due to the shortage of workers and the necessity to employ lower skilled people. Williamson (2014) notices that historically, unemployment conditions in the destination country always dominate, therefore migration as a result of global recessions tends to worsen the situation in the origin country while it softens the conditions in the destination country.

From these theories we can obtain another key conclusion, that is, that unemployment increases migration.

Unemployment can also be included in the *Net return* formula seen above, because there is a risk when individuals migrate as they might not find a job immediately, or they find jobs for which they are overqualified and earn very little (Bodvarsson et al., 2014):

$$Net\ return = \sum_{t=1}^{T} \frac{(p(t)E_t^d - E_t^{Sp})}{(1+r)^t} - \sum_{t=1}^{T} \frac{(C_t^d - C_t^{Sp})}{(1+r)^t} \quad (2.2)$$

Here "p(t)" is the probability that a migrant will be employed in period t. Most of the literature following Sjaastad (1962) is just an extension of his model, usually adding motives for migration or varying migration costs. For example, the new economy of migration (Moreno Torres and López Casasnovas, 2006), states that the decision to migrate is taken within the family unit, not individually, to diversify risk and avoid credit constraints. If some family members are abroad, they can transfer money to the poorer country.

The most important theory developed after Sjaastad and that is key for this paper is the relative deprivation theory. The main idea is that when a family sends one of its members to work abroad, it does not do it only to improve income in absolute terms, but to improve its relative position with respect to a reference group (Moreno Torres and López Casasnovas, 2006). An increase of income in the origin country may increase migration if inequality increases. This theory was first developed by Stark (1984). Although he only focused on least developed countries, Borjas (1987) applied it to developed countries as well. Therefore it is an important part of the analysis of this paper.

Stark (1984) built his theory upon the observation that migration from rural to urban areas was higher from villages where income distributions were more unequal, and migrants were predominantly the poorest. His approach is based on the assumption that absolute income changes do not play a role in migration, but relative deprivation does. From this point, simple static models with exogenous wages started to evolve to migration as a dynamic decision and endogenous wages. I will focus more on Borjas' (1987) approach to this issue, which was extended later by Clark et al. (2007) and Mayda (2010). Continuing the direct application of the theory to the case of Spain, if we restrict our study to two countries *i*, the origin country Spain (Sp) and destination (d), we can model earnings *w* distributed as:

$$\ln w_i = \mu_i + \varepsilon_i \quad ; \quad \varepsilon_i \sim N(0, \sigma_i^2) \quad ; \quad i = Sp, d \tag{2.3}$$

Earnings always refer to those expected by an individual that can either stay in Spain or migrate, with μ_d referring both to the mean income Spanish migrants would earn if all Spanish population moved to a destination d and the mean income of natives from destination d. This is known as a Roy model. Furthermore, it is assumed that errors ε_{Sp} and ε_d have a correlation coefficient ρ. From this, Borjas obtains an Index function (I) that stands for the migration decision in the origin country:

$$I = \left(\mu_d - \mu_{Sp} - \frac{c}{w_{Sp}}\right) + (\varepsilon_d - \varepsilon_{Sp}) \quad (2.4)$$

where *c* stands for costs of emigrating. When this Index function is positive, then there is emigration from Spain. Therefore, Borjas derives from it the emigration rate as a probability *P*, which is a negative function of mean income in Spain and costs of emigrating and a positive function of mean income in the destination (contrary to the beliefs of Stark, who assumed in his model that income does not affect migration):

$$P = Pr\left(v > -(\mu_d - \mu_{Sp} - \frac{c}{w_{Sp}})\right); \; v = (\varepsilon_d - \varepsilon_{Sp}) \quad (2.5)$$

Relative deprivation and skill selection start playing a role when comparing the standard deviations of the errors with the correlation ρ. When ρ is higher than $\frac{\sigma_{Sp}}{\sigma_d}$, then the average emigrant is better off in terms of earnings than the average person in Spain. We can also compare it from an immigrant at destination point of view. Borjas then examines the income differential between the emigrant and the average person in the origin (Spain) and in the destination. If it is positive for Spain and destination d, which happens when ρ is high and inequality is higher in destination, then we observe positive selection, that is, people from the upper tail of the distribution in Spain will emigrate and they will outperform natives from destination, will improving their relative position in the distribution. If both are negative, inequality is higher in Spain and those at the bottom of the distribution will emigrate (but they do not outperform natives at destination), which is also the case when only the income differential in Spain is negative (although here migrants outperform natives).

In short, when there is more inequality in the destination country ($\sigma_d^2 > \sigma_{Sp}^2$) there will be a brain drain situation (high skilled and higher income individuals will migrate), while if there is more inequality at home, low-income workers will be the ones who leave the country. Clark et al. (2007) reach this same conclusion, but they divide costs in constant and individual specific costs, and wages depend on a skill premium.

This research was expanded by Stark and Taylor (1991), who theorize that when human capital is not valued in the destination country, and immigrants can only get jobs that only require low skills, higher education in a country will decrease its emigration rates.

Nevertheless, the conclusion of Borjas' model can be contradicting, because if the gain is high enough, migration can take place from rich to poor countries. But the hypothesis of relative deprivation has some key ideas, such as the importance of the income distribution and the not so important absolute income differences (Bodvarsson et al., 2014).

From the relative deprivation theory we reach another conclusion: higher inequality in the origin country will increase migration with respect to countries where inequality is lower.

As for the role of welfare in migration, government transfers as a form of social insurance can substitute earnings while migrants search for a job. Therefore, welfare benefits attract migrants (Bodvarsson et al., 2014).

Going back to Moreno Torres and López Casasnovas (2006), they describe the world system theory as migrations that take place due to capitalist economic relations starting in less developed countries (globalisation). The flow of workers follows the flow of goods and capital in the opposite direction, especially when it was a colony or both countries have similar cultures and languages.

Finally, it is important to address the importance of migration persistence through the migration networks and institutional theories. According to them, migration is independent of push and pull factors, and the existence of migration networks (family connections, same nationality…) increases the probability of migrations and reduces migration costs. As this network grows, migration flows increase, and organisations are created and they procure transport, work, etc. (Moreno Torres and López Casasnovas, 2006). Bodvarsson et al. (2014) compile a range of theories, from the idea that migration to an area increases if migrants already have friends of family there, languages are similar or there is a large migrant stock in the destination country. This can in fact be included in the theoretical models as part of the costs in the form of a risk variable that is lower when migrant networks grow.

2.2. Empirical literature

In this section I will outline the main empirical findings in the literature regarding international migration and migration in Spain, focusing especially in the application of the theoretical models detailed previously.

Most of the literature on international migration focuses on immigration, although there is also research on emigration and its impact on the source countries, although a majority of it is related to Mexican emigration to the United States and how it has affected wages in Mexico, and how emigration rates vary significantly across regions and education groups. In fact, there is evidence on emigration raising wages in regions with higher emigration rates. There is also research on how changes in economic regimes influence migration, such as the transition from Communism in the east of Europe (Blau and Kahn, 2014).

Focusing first on the broad empirical research related to international migration, most of it refers to the topic of the influence of wages, income or GDP per capita on migration. For example, Son and Noja (2012) study the neoclassical theory, and they find that income, wage differentials and worse living conditions (measured with GDP per capita) decrease emigration flows to the EU. Beine et al. (2013) use a gravity model with data from 30 countries between 1980 and 2010 to test the influence of business cycles, and they find that an increase in the wage ratio increases bilateral migration flows, while the expected wage is affected by the cyclical stance, which is a signal of future probability of employment. Carlos' (2002) empirical analysis on Philippines' emigration from 1981 to 1995 relies on fixed effects panel data to calculate the negative influence of average domestic earnings when focusing on push effects, while for pull effects the signs are the opposite. Enache and Pânzaru (2012) study migration flows from Romania to 12 European countries which have a higher income per capita, and they find evidence of the positive influence of GDP per capita. Mihi-Ramírez et al. (2014) focus on migration in the European Union and how it is influenced by earnings, and they find the same sign as in the rest of the literature. Sulaimanova and Bostan (2014) research the determinants of emigration using a gravity model, focusing on flows from Tajikistan and Kyrgyzstan between 1998 and 2011. Their findings are in line with the rest of the literature, and they include the effects of GDP per capita and wages.

Different approaches include Hanson (2010), who reviews different studies that find an inverted U relationship between average income (GDP per capita) in the origin and emigration. Tilly (2011) focuses on migrant stocks, and concludes that they have a negative relation with changes in income per capita in the destination countries. Finally, Mayda (2010) researches what affects international bilateral migration flows using a gravity model. Her main findings include the positive effects of pull factors, measured as increased income opportunities at the destination, and the scarce effect of push factors, with both of them being heavily influenced by migration policy.

Another big topic in the empirical literature is the influence of unemployment. Son and Noja (2012) argue that, in line with the macroeconomic side of the neoclassical theory, labour markets are key to understand migration flows. For example, Beine et al. (2013) include in their gravity model relative employment rates and find a negative impact of unemployment, same as Merkus (2015), although his analysis has been applied to internal migration. Also related to business cycles and employment is Bodvarsson et al. (2014), that reviews a special sensitivity of migration when changes in employment conditions take place in the destination, while Carlos (2002) concludes that employment rates are more important for Philippines international emigration than earnings. In their analysis on Romanian emigration flows, Enache and Pânzaru (2012) find evidence of the positive influence of the employment rate in the destination country, same as Mihi-Ramírez et al. (2014).

The other key subject in the literature since the publication of the theory developed by Stark (1984) is the role of relative deprivation in the determination of migration flows. He applied his theory to rural and urban migration flows, and found that the largest migration flows are not found in the poorest regions, but in the most unequal. Since then, there has been an extensive amount of empirical research on relative deprivation, starting with Borjas (1987), who also tested his theory but using immigration flows from 41 countries to the US from 1951 to 1980. He finds that migrants from more developed countries increased their earnings relative to their skills, while the opposite happened for those coming from poor countries. On top of that, political and economic conditions at the origin countries explain the majority of the variations in income for the immigrants with the same skills: those with higher incomes came from countries with higher GNP, lower inequality and "politically competitive systems".

However, while Liebig and Sousa-Poza (2004) prove empirically than human capital has a positive effect on migration, in contrast to Borjas' model and in line with other research, they conclude that positive self-selection happens even in countries with high levels of inequality if high-skilled workers have lower migration costs. Therefore, they argue that income inequality influence on emigration is not as important, contrary to Borjas. Czaika and de Haas (2012) expand Borjas' analysis and find proof that when absolute deprivation decreases, emigration increases. For possible migrants, though, relative deprivation measured with respect to the rest of the world is not as important compared to when it is measured with respect to populations in other countries. Inequality can be a signal of social mobility and opportunities to improve everyone's situation, so when a country has inequality, they find that it attracts migrants. Son and Noja (2012) also find evidence on the influence of inequality of emigration.

Hanson (2010) also includes income inequality, and concludes that when the effect on migration is positive, migrants are negatively selected in terms of skills. Felbermayr et al. (2014) review a series of studies that research how, even when skills are not observable, positive selection can still take place. This happens because only those with higher skills have an incentive to migrate when it is costly to do so, and therefore will reveal their skills through migration. This is important for origin countries, who could experience brain drain, and destination countries, as migration flows affect income distribution. Finally, Clark et al. (2007) find non-linear effects of inequality, same as Mayda (2010).

As for the research related to the importance of welfare and education, Belot and Hatton (2012) focus their research on highly educated immigrants from 80 source countries in 29 OECD countries. They develop a variation of the Roy model to estimate what determines educational selectivity both at home and at the destination and find that a higher relative return to skills favours positive selection regarding skill level, although only when origin countries are constrained due to distance (cost of migration) and poverty (liquidity constraint). Other research includes Enache and Pânzaru (2012), who conclude that public expenditure on family and children in the destination country attracts Romanian migrants and that there is a positive influence of life expectancy in the destination country.

Similarities and relations between countries are also important when studying migration. Although Belot and Hatton (2012) find that cultural differences do not have a

clear effect, as sharing a language increases educational selectivity but also allows low-skill individuals to fit in the destination country, Beine et al. (2013) find that agreements like Schengen, in the case of Europe, significantly increase international mobility of workers. This has happened since the 1990s due to the lower migration costs, which is good when facing asymmetric shocks. Ullah (2012) also researches international migration from the perspective of the origin country (Bangladesh) using a gravity model, and although he finds that economic factors are key, empirics show some cultural similarities such as language are also important. In fact, studies show the importance of language for migrants, which can affect employment opportunities in the destination country. Chiswick and Miller (2014) include as a cost the acquisition of the language, and it is lower the more similar the languages are. Other links include former colonies, which favour the size of migration (Czaika and de Haas, 2012).

Regarding the importance of migration networks, Clark et al. (2007) find a positive influence of the stock of previous migrants. Other topics include the negative influence of inflation on emigration flows (Son and Noja, 2012), that distance is negatively associated with migration (Clark et al., 2007; Mayda, 2010; Czaika and de Haas, 2012) and the positive influence of population growth on the probability to migrate (Carlos, 2002).

Moving on to the characteristics of migration in Spain, studies that refer to other countries also include Spain's emigration, such as Bertoli et al. (2013), who analyse how the last crisis has affected migration from 28 EEA countries to Germany using data from 2006 to 2012, and including not only bilateral factors, but also how changing conditions in other possible destinations affects migration to Germany as well. They reference the situation in Spain, where the labour market conditions and confidence in government bonds deteriorated and migration flows to Germany increased at the same time. More in general, economic conditions at the origin and alternative countries affect the scale of bilateral migration flows, although common factors reduce them. Their analysis finds that intra-European migration flows are very sensitive to these alternative destinations. Jauer et al. (2014) study how migration relates to unemployment in regions of the European Union and the United States from 2005 to 2011, although they only measure net migration and do not take into account the expectations on future earnings. They find strong push factors in the emigration from the Southern European periphery and Ireland due to the increasing economic gaps within the Eurozone.

As most of the research for Spain's migration situation has focused on immigration rather than emigration, it would be wise to include some results here, as the variables can also be applied for this study. Moreno Torres and López Casasnovas (2006) study the drivers of immigration to Spain from 1994 to 2004. They find that networks effects, differences in unemployment rates, distance and cultural similarities, among other factors, are key to determine these flows. Márquez Arboleda et al. (2004) also study immigration to Spain from the 1990s using panel data. Among their findings are the negative effect of GDP per capita in the origin country (in line with neoclassical theory and gravity models), corruption and distance; the positive effect of unemployment in the origin country (in line with Keynesian theory), migrant stock (in line with network and institutional theory), Spain's population size (in line with gravity models) and cultural similarities. Finally, Bertoli and Fernández-Huertas Moraga (2011) focus on immigration flows from 61 countries to Spain from 1997 to 2009, and instead of following Borjas' (1987) theoretical model, they use a Random Utility Maximization model which is based on:

$$U_{ijk} = V_{ijk} + \varepsilon_{ijk} \quad ; \qquad k\text{=individual; } i\text{=origin; } j\text{=destination} \tag{2.6}$$

where U refers to utility and V can refer to deterministic terms such as wage, as well as include costs of migrating and location-specific terms. They also introduce a common correlated effects estimator developed by Pesaran to account for multilateral resistance to migration, that is, the attractiveness of other destinations when making the decision to migrate [(Mayda (2010) also accounts for this problem but adding as a variable the average of GDP per worker of other destinations]. They find that an increase in real GDP per capita reduces migration rates to Spain.

As for the studies that do focus on emigration, or include it along with immigration, many of them have used data that corresponds to the 1960s. More recent studies have been developed only after the economic crisis hit Spain, and include Izquierdo et al. (2014) who study Spain's emigration and immigration flows from 2008 to 2013 and find a positive relationship between unemployment and emigration, and that differences in emigration rates for different nationalities can be explained through emigration costs. These costs decrease when the emigrant stock increases abroad. As for skills, those with higher education levels have a higher probability of emigrating. This brain drain could have effects on Spain's potential growth in the future. They expanded this paper later on,

but still focusing mainly on how unemployment affects migration flows in Spain. They find that the composition and scale of emigration has changed. Both nationals and non-nationals respond in a similar way to unemployment changes, but migration networks are not very important for emigrants born in Spain, although they could develop in the future. They also observe positive selection on education for both groups, although since 2008 the number of less skilled migrants has increased (Izquierdo et al., 2015).

Ramos and Royuela (2016) study emigration rates from Spain from 2008 to 2013 (like Izquierdo et al., 2014 and 2015), although only for graduates. They highlight that, although nationals' emigration rates have increased, most of Spain's emigrants are people who were not born in the country and are going back to where they are from. They test from a macro perspective if there has been positive selection in Spain's emigration, that is, graduates have had a higher propensity to migrate. Among their findings is that while income differentials play a role in Spanish migration, unemployment differentials do not, although the crisis has been an important push factor. From a micro perspective, they find evidence of positive selection and that language impacts migration decisions. Finally, Rodríguez-Fariñas et al. (2016) study the causes of Spanish emigration to Mexico due to the last economic crisis. They find that generally, emigrants are young people highly qualified that have been attracted to Mexico due to the work opportunities and the old migration networks.

2.3. Hypotheses

The hypotheses that I will test later on are based on the theoretical framework, the empirical literature and Spain's current situation that was explained in the introduction. The hypotheses are divided in four blocks: the benchmark model, the main hypotheses of the augmented model, the welfare variables and the relations between countries. All the hypothesis will be tested for the two groups of emigrants (nationals and non-nationals) and for different destination regions (European Union, rest of Europe and North Africa and South America). A list of the 60 countries and their regions can be found in Appendix A. A separate robust check will be also done for a subset of the data from 2006, due to the discrepancies that are explained in the emigration part of Section 3.1.

As for the benchmark model, the main variables included, according to the gravity model explained later in Section 3.2, are distance and GDP per capita:

- H1: an increase in GDP per capita in the origin (destination) decreases (increases) emigration flows from Spain and distance has a negative influence on these flows.

This follows from the neoclassical theory outlined previously, with GDP per capita as a proxy for wages or earnings, following also the empirical literature such as Márquez Arboleda et al. (2004) and Enache and Pânzaru (2012). The hypothesis of distance results from the theory and empirics that point out that migration costs are key to the analysis. Belot and Hatton (2012) include distance as part of the cost of migration, while Clark et al. (2007) find that it negatively influences migration size.

If we expand the benchmark model slightly by including other migration costs and basic variables, the following hypotheses can be formulated:

- H2: costs of migrating such as different language and higher inflation in the origin country decrease emigration flows. Bigger population in both countries increases emigration flows.

Following Lewer and Van den Berg (2008), the interaction term of both populations has a positive relation with migration flows. As for inflation, it follows from Son and Noja (2012), while language follows from Bodvarsson et al. (2014), who find that cultural similarities, including language, increase migration flows, while Chiswick and Miller (2014) find that sharing a language affects the employability of migrants.

Moving on to the augmented model, its main pillars are uncertainty and relative deprivation. The first one is proxied by unemployment rates, since as seen in the literature review, it is a risk that migrants face both at home and abroad, at can also come across as a cost when moving to the destination country. As for relative deprivation, it is measured by inequality.

- H3: higher unemployment in the origin (destination) country increases (decreases) emigration flows from Spain, while higher relative inequality in the origin (destination) country decreases (increases) emigration flows from Spain.

The first part of this hypothesis reflects mainly the theoretical and empirical work that has followed Sjaastad's work, as can be seen in Izquierdo et al. (2015). The relative deprivation part is built from the theory developed by Stark (1984) and Borjas (1987). As

this has not been applied to the Spanish case, I follow the empirical literature on international migration, such as Mayda (2010) and Clark et al. (2007). As Spain has already medium-high levels of inequality, compared especially to other rich countries in the dataset, and these levels have increased since 2002, I follow Mayda's (2010) result, consistent with the theory of Borjas (1987), that in countries with already high inequality, this relation is negative. However, if this is not linear (as in Graph 3.1), it is necessary to include also the quadratic form, which would have the opposite sign.

Regarding the welfare variables, the hypotheses are divided in two, one referring to variables measured in years, and the other for variables referring to percentages of GDP.

- H4: higher levels of education in the origin country will increase emigration flows from Spain and improving living conditions (life expectancy) in the origin country decrease emigration flows.

As there is no available data for the level of education of emigrants (see Section 3.2), this will be measured using educational attainment data referring to at least post-secondary completed level. This will be a proxy for level of skill in Spain and the destination countries. An increase in education in Spain should increase emigration rates if we follow the study of Izquierdo et al. (2014), and assume that we observe positive selection and brain drain. If there is negative selection, then the relationship would be negative (which would be in line with Stark and Taylor, 1991).

In the case of living conditions, I follow mainly the empirical work of Enache and Pânzaru (2012), who find that life expectancy in the destination country has a positive effect on emigration.

- H5: higher relative taxes and health and education expenditures in the origin (destination) country decrease (increase) emigration flows from Spain.

This hypothesis follows Bodvarsson et al. (2014) and Enache and Pânzaru (2012), who find that welfare benefits attract migrants. In the case of Spain, this would be the case especially for those countries that offer a stronger welfare system compared to the Spanish one, as it can be a form of social insurance. Taxes are a proxy for state intervention, following Enache and Pânzaru (2012).

Finally, the last block of hypotheses refers to the relations between countries:

- H6: a bigger Spanish migrant stock and higher FDI inflows in the destination country, as well as more trade between Spain and the destination, increase migration flows.

For this I rely mainly on the work of Bodvarsson et al. (2014), Moreno Torres and López Casasnovas (2006) and Clark et al. (2007), as well as on the extensive theoretical literature of migration networks, that finds a positive relation between migrant stock and migration flows. The world system theory is the base for this hypothesis (Moreno Torres and López Casasnovas, 2006), as globalisation, or more trade and capital flows between countries. Bodvarsson et al. (2014) find that cultural similarities, including language, increase migration flows, while Chiswick and Miller (2014) find that sharing a language affects the employability of migrants.

3. Data and Methodology

This section is divided in two sub-categories: first I will explain the variables that are part of the model and the main sources of the data, as well as the limitations that I encounter when gathering it and that, therefore, have an impact on the results; finally, I will review the model that will be used to test the hypotheses and that will be based on the gravity model of international migration.

3.1. Data

In this section I will explain the data that will be used in the model and the main drawbacks of some variables, focusing especially on emigration data. Furthermore, I will describe the situation of unemployment and inequality in Spain, as well as the regional characteristics of these three variables that, although not included in the models, are still important to understand the phenomenon of emigration in the country.

In Table 3.1 a list of all the variables and their descriptions, classified in the same blocks as the hypotheses in Section 2.3, can be seen. I will explain the variables divided in two groups: Section 3.1.1 refers to emigration, the dependent variable, while Section 3.1.2 refers to the explanatory variables. All the variables refer to years 2002 to 2013. The full list of variables with their sources can be seen in Tables A.2 and A.3 in Appendix A, while the descriptive statistics can be found in Table A.4 in the same Appendix.

Table 3.1 - List of variables and description

Block	Hypothesis	Variable	Description
Dep.	Dep.	Log(Emigration)	Logarithm of total Emigration flows from Spain to known destination countries.
		Log(Emigration of nationals)	Logarithm of Emigration flows of nationals from Spain to known destination countries.
		Log(Emigration of non-nationals)	Logarithm of Emigration flows of non-nationals from Spain to known destination countries.
1	H1	Log(Relative GDP pc)	Logarithm of ratio of GDP per capita in Spain over GDP per capita in destination country (constant US$).
		Log(Distance)	Logarithm of Distance (in km) between Madrid and capitals of destination countries.
	H2	Log(Population Spain * Population destination)	Logarithm of interaction term Population in Spain * Population in destination country (in thousands).
		Relative Inflation	Ratio of Inflation in Spain over Inflation in destination country (measured as GDP deflator, in %).
		Language	Dummy variable with value 1 for Spanish, 0 for the rest.
2	H3	Relative Unemployment	Ratio of Unemployment in Spain over Unemployment in destination country (measured over share of labor force).
		Relative Youth Unemployment	Ratio of Youth Unemployment Spain over Youth Unemployment in destination country (measured as Share of youth not in education, employment, or training over total of youth population).
		Relative Inequality (Gini or S80/S20)	Ratio of Gini or S80/S20 index in Spain over Gini or S80/S20 index in destination country (Gini ranges from 1 to 100).
		Square of Relative Inequality (Gini or S80/S20)	Square of ratio of Gini or S80/S20 index in Spain over Gini or S80/S20 index in destination country (Gini ranges from 1 to 100).
3	H4	Relative Educational Attainment	Ratio of Educational attainment in Spain over Educational attainment destination country (measured as % of population over 25 years old with least post-secondary education completed).
		Relative Life Expectancy	Ratio of Life expectancy in Spain over Life expectancy in destination country.
	H5	Relative Expenditure on Education	Ratio of Government expenditure on education in Spain over Government expenditure on education in destination country (measured as % of GDP).
		Relative Expenditure on Health	Ratio of Health expenditure per capita in Spain over Health expenditure per capita in country (measured in PPP, constant 2011 international $).
		Relative Taxes	Ratio of Tax revenue in Spain over Tax revenue in destination country (measured as % of GDP).
4	H6	Log(Migrant stock)	Logarithm of stock of Spanish emigrants in destination country.
		Relative FDI	Ratio of FDI inflows in Spain over FDI inflows in destination country (nominal million $).
		Trade	Ratio of Trade of destination country with Spain over World Trade of Spain (measured as Exports + Imports over GDP).
5	Dummies	Europe, Africa, EU, America	Dummy variables for regions.

Notes: (i) subscript "Sp" indicates Spain; subscript "d" indicates destination country; (ii) subscript "t" indicates year.

3.1.1. The dependent variable: Emigration

The main variable for this analysis is emigration. I have chosen gross emigration flows instead of net migration because this thesis focuses only on the recent phenomenon of high emigration from Spain, although there is a wide range of literature that uses either net migration or immigration.

The data on emigration is taken from Instituto Nacional de Estadística (Nationals Statistics Institute; INE) because it offers a longer timespan than any other source, such as Eurostat. The data goes from 2002 to 2013 and can be divided into nationals and non-nationals. INE offers two different datasets for emigration: *Estadística de Variaciones Residenciales* (Residential Variation Statistics or EVR) and *Estadística de Migraciones* (Migration Statistics or EM). As Izquierdo et al. (2015) point out, until 1998 there was no organised way to keep record of migration flows. EVR was only made available in 2002, and the data for the second source, although much more complete, it is only provided by INE from 2008, therefore I have decided to focus only on the first source following Domingo i Valls et al. (2014). It is important to know that if emigration data is obtained by destination country, there is no information available regarding province of origin, education, sex or age, only nationality. Knowing the education levels would be useful to differentiate between high skill and low skill individuals who do not respond in the same way to income inequality differentials.

Municipal registers are the main source for emigration flows contained in EVR, but this data has to be handled with caution. Although outflows of Spanish nationals should be more reliable than those of non-nationals, it relies on information from consulates and embassies. As only permanent emigrants have incentives to register, and they do not always do so, this data may not be completely accurate (Izquierdo et al., 2015).

Domingo i Valls et al. (2014) elaborate also on the reasons why emigration data is incomplete. Due to the high levels of immigration that Spain experienced since the 1990s, the INE and the legislation were focused on registering everyone efficiently at each municipality. However, emigrant data is of worse quality because of various reasons. First, emigrants do not have any incentives do deregister, and up until 2006 if they left, they remained registered, even if they were immigrants that returned to their countries (INE also warns of this). Second, Spanish citizens tend to not register in the consulates and embassies. In both cases, while being registered has benefits such as the public health

system or public education, deregistering doesn't have any. Third, municipalities do not want to see the number of inhabitants fall, as their finances depend on it. In an effort to fix these deficiencies, from 2004 a new type of deregister, the cancellations for undue inscription, and from 2006 the cancellations by expiration. Non-nationals are the ones that have been especially affected by this measure, as can be observed in table 3.2. While the data I collect is not 100% of all emigration flows, as not all regions are included, it only refers to those emigrants of whom we know their country of next residence. Therefore, all those who are deregistered using the new measures, plus those who do not say where they go next, cannot be included in the EVR.

Table 3.2 – Percentage of total emigration flows included in the analysis

NATIONALITY	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
NON-NATIONALS	90,74	90,78	30,00	34,09	18,41	14,19	15,50	13,00	10,87	11,85	18,18	17,44
NATIONALS	74,43	76,04	76,35	85,71	82,94	80,84	81,13	81,84	83,22	83,35	85,71	86,00

Notes: (i) Total flows according to Estadística de Variaciones Residenciales; (ii) percentage of emigrants only with known destination country in Europe, Latin America and north of Africa; (iii) discrepancies in the percentage of non-nationals due to registration laws
Source: INE

Domingo et al. (2012) point out that for almost half of foreign (either born abroad or non-nationals) the information of destination country is not available. This could result in some inconsistency in the results, therefore, as nationals' data is more complete, the analysis will be carried out for both groups together and separately.

Another source for migrant stock, which could also be used to measure migration flows, is the *Padrón de Españoles Residentes en el Extranjero* or PERE (census of Spanish people with residence abroad). It also elaborated by INE, but Domingo i Valls et al. (2014) have also preferred to use the EVR due some problems with this census. For example, migration flows cannot be measured directly, as it includes also births and deaths. It also includes those who obtain the Spanish nationality, even if they have never been in Spain. This is because a Law was passed in 2007 that recognises the descendants of those who had to migrate unwillingly between 1936 and 1955 (the period of the civil war and the start of Francos' regime).

Once the issue with the quality of the data is dealt with, it is important to describe generally these emigration flows. The 60 countries chosen for this analysis of Spanish

emigration flows can be found in the analysis. I have limited this analysis to countries in Europe, Morocco and Algeria, as they are either in the same continent or really close and have an established immigrant network in Spain, as well and Latin American countries, since most of them were previous colonies of Spain and share the same language. As can be seen in Table 3.3, there are differences in the destinations of emigrants depending on their nationality. For non-nationals, their main destination is their own origin country, as the highest percentages of emigrants correspond to the biggest immigrants' communities in Spain. Therefore, the main destinations are, apart from the biggest countries that did not suffer from the crisis as much as Spain, Bulgaria, Portugal, Romania, Ukraine, Morocco, Colombia, Ecuador or Bolivia. For nationals, however, France, Germany and the United Kingdom are the main destinations (probably due to the facilities offered by the EU and Schengen), although some Latin American countries and Switzerland are also important, probably due to the language similarities (in the case of America) and old migration networks.

Table 3.3 - Emigrants by nationality and destination (% of total 2002-2013)

Country	Nationals	Non-nationals	Region	Country	Nationals	Non-nationals	Region
Belgium	**4,35%**	1,35%	EU	**Andorra**	**2,87%**	0,21%	Europe
Bulgaria	0,13%	**2,63%**	EU	**Switzerland**	**4,73%**	0,80%	Europe
France	**11,92%**	3,31%	EU	**Ukraine**	0,06%	**2,63%**	Europe
Germany	**8,85%**	5,21%	EU	**Morocco**	2,21%	**7,11%**	Africa
Italy	**2,92%**	2,09%	EU	**Argentina**	**5,52%**	3,77%	America
Netherlands	**2,22%**	**2,05%**	EU	**Brazil**	2,40%	**3,19%**	America
Portugal	1,78%	**3,14%**	EU	**Colombia**	2,48%	**4,36%**	America
Romania	0,23%	**11,87%**	EU	**Ecuador**	**6,93%**	**6,35%**	America
United Kingdom	**14,62%**	**12,95%**	EU	**Peru**	1,63%	**2,16%**	America
				Venezuela	**4,74%**	1,68%	America
				Bolivia	1,20%	**5,55%**	America
				Chile	**2,03%**	1,55%	America

Source: INE

Note: the percentages correspond to each category separately (100% of nationals, 344,424 emigrants, and 100% of non-nationals, 367,333 emigrants).

Finally, although due to the limitations of EVR, regional migration data cannot be included in this analysis, it is important to specify that not all Spanish regions[6] register the same outflows (absolute or relative). Almost 40% of emigrants come from Madrid

[6] A Map can be found in Appendix B.

and Barcelona, and many of the regions that are the main destination for both internal migration and immigration are now the main emigrant regions. This has derived in high emigration flows from the Mediterranean area and the islands, which attracted immigrants in construction, agriculture and tourism (Domingo i Valls et al., 2014).

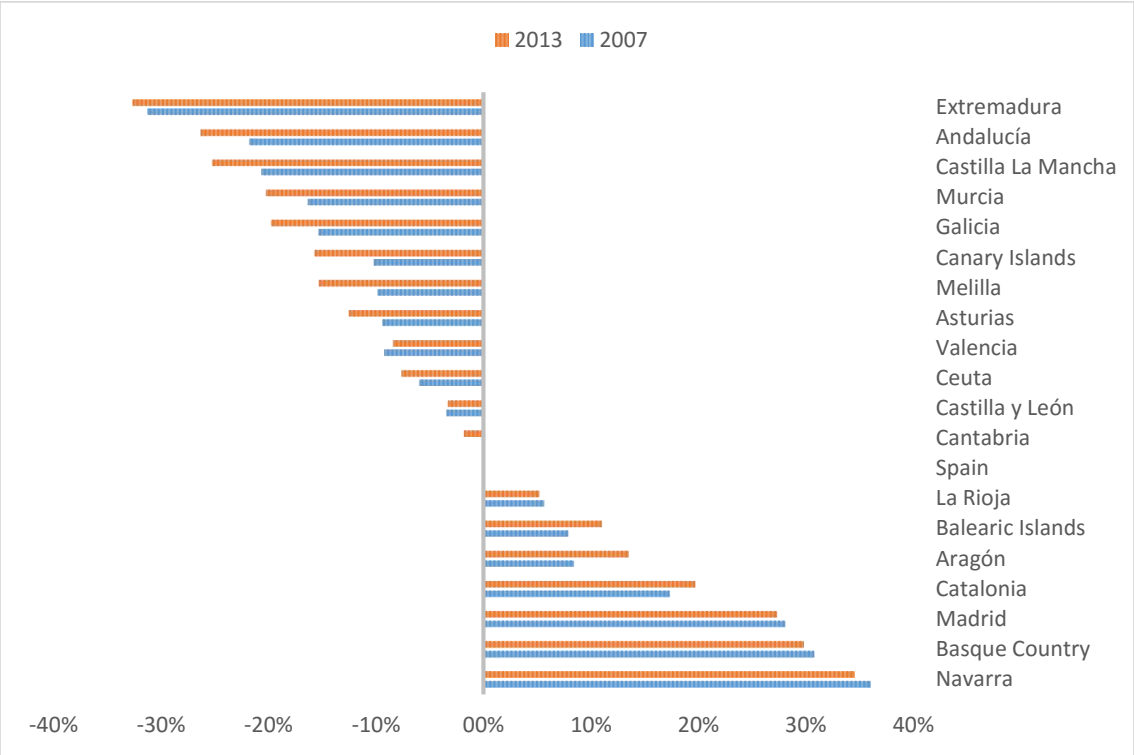
3.1.2. The explanatory variables

As for the rest of the variables that will be used in the analysis, I will again divide them in groups: benchmark model, augmented model, welfare and relations. The variables that have one value for Spain and another for the destination country will be used as ratios of Spain over destination in the models (except for population).

Benchmark model

The first variable that is part of the benchmark gravity model (which will be explained in Section 3.2) is the effect of income or earnings on emigration rates. As most of the literature uses GDP per capita as a proxy for income, I will also use the same variable. GDP per capita is obtained from the World Bank Database and measured in constant US$, as is the international currency. The data is for Spain and the 60 destination countries. Although the evolution of GDP per capita was explained already in the Introduction, it also important to acknowledge that regional differences in Spain can be very high, with deviations of up to 30% with respect to national levels, as can be observed below in Graph 3.1.

Graph 3.1 – Comparative of GDP pc by regions with respect to national GDP pc



Source: INE

Southern regions such as Extremadura or Andalucía register GDP per capita levels of more than 25% below national figures, while the traditionally richer regions like Madrid, Catalonia, Basque Country and Navarra rise above 20%. By using only national data on both GDP and emigration we are treating all regions the same, even though their situations differ greatly. Therefore, it is important to know that these disparities exist and that could lead to very different results if the data was available.

Continuing with the benchmark model, population and distance are also included, with the first being obtained from the World Bank. Both of them are basic variables in the gravity models, with distance being a proxy for migration costs. It refers to the kilometres between the capitals of Spain and the 60 destination countries.

Finally, inflation (as a deflator of GDP) is also included in the benchmark model as another proxy for migrations costs: when inflation is higher in the destination country, it would be costly to move there, therefore being a deterrent of emigration from Spain. Language is also a form of cost, and is used as a dummy variable that has the value 1 when Spanish in an official language in the destination country, and 0 otherwise.

Augmented model

In the augmented gravity model that will also be explained later I will include those variables that because of Spain's situation and the importance in the literature are of special interest for this thesis. The first one is unemployment, of which the evolution in Spain can be seen in Table 3.4 and that was obtained from the World Bank. This table also includes the share of youth (from 15 to 29 years old) that is not in education, employment or training. Again, these variables refer to Spain and the destination countries.

Table 3.4 – Unemployment in Spain (% of active population/share of youth)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
UNEMPLOYMENT	11.6	11.5	11.2	9.3	8.6	8.4	11.5	18.1	20.2	21.7	25.2	26.3
YOUTH NEITHER STUDYING NOR WORKING	-	11.7	11.8	13.0	11.8	12.0	14.3	18.1	17.8	18.3	18.6	18.6

Source: World Bank

As can be seen above, unemployment is a chronic problem for Spain. There have been numerous reforms of the labour market over the years, for example in 2001, when the unemployment rate was 10.5% and had fallen from around 20% in 5 years. As the previous reforms had worked, this new one focused on expanding what was thought were the strongest developments implemented: favour par time jobs and transform temporary jobs in indefinite. Over the course of the next four years, both type of contracts increased significantly, and unemployment fell below 10%. However, temporary contracts were still over 30% of the labour market, the highest in the EU, which motivated the government to pass new legislation in order to reduce this rate. Again the objective was to promote indefinite contract's growth, while trying to maintain protection to unemployed people and increase the occupancy rate. Until the crisis hit, indefinite contracts increased, but temporary contracts maintained their cyclical phases, therefore perpetuating Spain's problems in the labour market (Gómez et al., 2018).

Despite the efforts of reform, once the crisis hit Spain unemployment rates soared, going from 11.5% in 2008 to 26.3% in 2013, along with other Southern European countries such as Greece, while most of the other developed countries have not reached 15%. This has led to more problems: the so called "lost generation" of young people and the long term unemployment, which makes reincorporating to the labour market more

difficult. Most of those who have been unemployed for over two years do not have secondary education, that is, they would be considered as low skilled workers. The high numbers are a result of the housing boom that took place before the crisis, which attracted young workers who gave up their education and that are now faced with incredibly high unemployment rates are no qualifications that would make it easier to find a new job (Jansen et al, 2016).

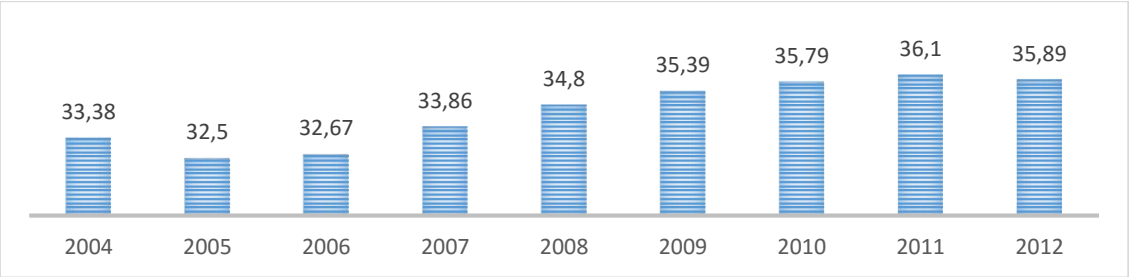
Again, new reforms were implemented in 2010 and 2012 to solve what have been called structural problems of the Spanish labour market. However, in the short term the effects cannot be observed, as unemployment has only begun to fall when GDP growth has improved. These reforms have also failed to improve the situation of young people under 25, whose unemployment rates are the most volatile, reaching almost 60% in 2016 (Jansen et al., 2016). Furthermore, around 20% of youth does neither work nor study (widely known as "ni-nis" in Spain), which makes their insertion in the labour market even harder.

Due to this situation in Spain, it would be expected to see high emigration rates to countries with lower unemployment figures, especially for young people. However, data by ages is not available for emigration, as was explained above. It can be seen, although not included in the analysis, the difference in unemployment by regions, which can be quite important. In fact, unemployment rates are clustered by regions in Spain. While high unemployment concentrates in the South-West, it is lower in the North, East and Madrid. These two groups of regions have had consistent differences of around 12-16 percentage points during the first decade of this century. Furthermore, there are differences in the individuals that live in those areas that affect directly the unemployment rates: families' composition and seize, and years of schooling (which can be seen as skill level), being these ones longer in the second group of regions (López-Bazo and Motellón, 2013). This can be directly related to the structure of the economy and the labour market in different regions, as those that depended more on the housing bubble were hit harder. The labour market relied more on construction workers (usually not skilled), who were later hurt more by the crisis and experienced higher unemployment rates (Fernández-Tabales and Cruz, 2013). The south of Spain is also traditionally an agricultural-dependent region, while most of the industry is located in Catalonia, Madrid and the Basque Country.

The next important variable refers to the hypothesis of relative deprivation, that is, the role of inequality on emigration. For this I will use two different measures: the Gini coefficient, which is broadly used in the literature (Mayda, 2010), and the S80/S20 coefficient, which is the one used for inequality in Eurostat. While Gini measures how the income distribution of the country deviates from a perfectly equal distribution and varies from 0 (equality) to 100 (although in the data used here is usually around 30-40), the S80/S20 ratio is obtained from the income distribution, dividing the income received by the top 20% of the population by the bottom 20%. The reason for using two different measures is that they are not perfect, as the Gini coefficient, for example, only gives a general view of income inequality, and the same coefficient can result from very different distributions (Jakobsson, 2006).

The Gini data for Spain can be observed below in Graph 3.2. It does not align with GDP growth, as it started to increase in 2006 when the Spanish economy was far from a crisis. This is because inequality is not only dependent on the economic situation of a country, but also by its characteristics (Ramos et al., 2015).

Graph 3.2 – Gini index in Spain (2004 – 2012)



Source: World Bank

As happened with other variables above, it is useful to compare the different Spanish regions to understand that inequality can vary a lot within a country. This would influence migration, but again, the data is not available. There are not many studies that focus on income distribution differences within Spain in the last decade, although Ramos et al. (2015) do it from 2003 to 2012. Instead of using the Gini index they prefer the Theil index, as its characteristics allows them to see differences of inequality between and within regions. They find that regions in the South such as Andalucia and Extremadura (which were the poorest), as well as the islands and Madrid (one of the richest) are very unequal, while others situated mainly in the North and the East are less unequal.

Furthermore, although inequality increased, these two groups have remained more or less stable after the crisis.

Welfare model

The variables that are part of the welfare model refer to education, health and government intervention.

As for education, it is measured in two different ways. First, educational attainment, which refers to the percentage of the population over 25 that has at least completed post-secondary education. It is a proxy for skill level and is obtained from UNESCO. And second, education expenditure of the government as a percentage of GDP. Also as a percentage of GDP is measured health expenditure. Both variables can be found in the World Bank database, as well as life expectancy, which is also included.

Finally, government intervention is measured as taxes as a percentage of GDP and is taken as well from the World Bank.

Relations model

The last extension of the model adds those variables related to relations between Spain and the destination countries. The first one is migrant stock as a proxy for migration networks, which is obtained from the Electoral Census (CERA), which is the most reliable source for Spanish citizens living abroad that covers every year and dates at least back to 2002. As it also happened with emigration flows, this variable has serious drawbacks: only permanent emigrants have incentives to register, and they do not always do so, therefore this data may not be completely accurate (Izquierdo et al., 2015). Furthermore, only people over 18 years old are part of this census, therefore a bigger part of the stock is not taken into account.

Other variables include FDI inflows (obtained from UNCTAD) as a proxy for the openness of the 61 countries, as well as trade with Spain (obtained from DATACOMEX and proxy for relations with Spain), which refers to the exports and imports over GDP between Spain and a destination country, divided by the total exports and imports over GDP of Spain.

Finally, there also 4 dummy variables for each of the regions (European Union, rest of Europe, Africa and America).

3.2. The gravity model

The literature focused on explaining international migration flows has mainly used gravity models when analysing data (Mayda, 2010; Ullah, 2012), although some authors have followed Random Utility Models and regressed emigration on unemployment (Izquierdo et al., 2015). As the gravity models are well accepted in the field (Anderson, 2011) and adapt perfectly to this study, allowing to incorporate bilateral variables in an augmented version, it is what I will use.

Gravity models are adapted from physics (Newton's Law of Gravitation), and what is known today as the benchmark gravity model was first applied in its most common form by Tinbergen (1962), and it is as follows:

$$Trade_{ij} = k \frac{GDP_i GDP_j}{Distance_{ij}} \quad (3.1)$$

where Trade = trade flows; i=origin; j=destination. Although gravity models in economics are more common in the field of international trade (Anderson, 2011), they can also be applied in international migration (Anderson and Van Wincoop, 2003; Lewer and Van den Berg, 2008). Gravity models especially fit the objective of this thesis as they allow to include both push and push factors that are discussed by Sjaastad (1962) or Moreno Torres and López Casasnovas (2006), especially because the empirical literature has found strong push factors for countries in the South of Europe (Jauer et al., 2014).

3.2.1. The benchmark model

As was explained in Section 2.1, Zipf (1946) developed the first stages of the gravity model when he related the volume of migration to the distance between two places and the product of the two populations. The key feature is the inverse relationship between the size of migration flows and the distance between the two countries, which as seen in the theory is considered a cost of migration (Sjaastad, 1962). However, instead of using population, the gravity models of migration follow those of trade and include the positive relationship between migration flows and the income of both countries.

To carry out the analysis the model is transformed into logarithms. Furthermore, when applying it to migration, Lewer and Van den Berg (2008) prefer to use a modified version, also in logarithms, that includes a term for relative income. Therefore, the gravity model

for international migration, when applied to our case of emigration from Spain, look as follows:

$$Log(Emigration)_{Spd,t} = \beta_0 + \beta_1 \text{Log}(Relative\ GDP\ pc)_{Spd,t} + \beta_2 \text{Log}(Distance)_{Spd,t} + \varepsilon_{Spd} \quad (3.2)$$

where "Sp" refers to Spain (the origin country) and "d" refers to a destination country. This form of the gravity model is in line with the theory of net return by Sjaastad (1964) that was explained in Section 2.1: people will migrate when the returns of doing so are positive, that is, when the difference in income between Spain and the destination country, measured as GDP per capita, is higher than the costs of moving, proxied as the distance between both countries.

In addition to this, the use of population is also justified as the more people in the source country, the more people could migrate, while the more people in the destination country, the bigger the labour market for migrants. The population term is an interaction of populations in Spain and the destination country, as used by Lewer and Van den Berg (2008). Some authors like Mayda (2010) and Clark et al. (2007) divide emigration flows by population instead of using an interaction term.

Before formulating the final benchmark model, it is important to acknowledge the time taken into account in the analysis: while some authors use one-year lagged independent variables (especially for GDP), like Mayda (2010), others prefer to avoid lagged variables as they judge that, although migration can be a result of expectations formed in the past, it can also be a result of a plan that was decided in case some major event, like losing a job, happens. This is the case of Jennissen (2004) and Ullah (2012). However, endogeneity and reverse causality must be addressed when using an econometric approach like this. Mayda (2010) argues that although migrants may influence wages, and therefore income, this relation can also be in the opposite direction, causing the estimates to be biased toward zero. She later discusses that reverse causality is not important in this type of analysis when using flows instead of stocks, but endogeneity is. To solve it, all variables that are time-varying are specified with one year lagged values.

Furthermore, as migration costs play a very important role in the decision to move to a different country, the benchmark model also includes language and inflation. Chiswick and Miller (2014) consider acquiring a different language as a cost, therefore reducing

migration, while Son and Noja (2012) find that inflation has a negative influence in receiving migration flows.

Therefore, if we add population, inflation and language to equation 3.2, which would cover hypotheses 1 and 2, the benchmark model (model 1 in the results) to be tested is:

$$Log(Emigration)_{Spd,t} = \beta_0 + \beta_1 \text{Log}(Relative\ GDP\ pc)_{Spd,,t-1} + \beta_2 Log(Distance)_{Spd,t-1} + \beta_3 Log(Population\ Spain_{t-1} * Population\ destination_{t-1}) + \beta_4 Relative\ inflation_{Spd,t-1} + \beta_5 Language + \delta_{Spd} + \varepsilon_{Spd,t} \quad (3.3)$$

where "t" ranges from 2002 to 2013, "d" refers to one of the 60 destination countries and δ_{Spd} are country fixed effects. Note that the emigration variable may refer to total emigration flows, emigration of nationals and emigration of non-nationals. Using logarithms may result in bias if the data contains zeros, however this is not the case here. All variables are bilateral (they apply to Spain and the destination country), because unilateral variables may result in biased models, although adding fixed effects should fix the problem (Lewer and Van den Berg, 2008).

The variable of GDP per capita, as all bilateral variables that I will use except for population, is in relative terms, that is, GDP per capita in Spain over GDP per capita in the destination country, instead of separating the logarithm. For this I follow Lewer and Van den Berg (2008), Mayda (2010) and Ullah (2012).

3.2.2. Extensions of the model: the augmented model, welfare and relations between countries

The extension of the gravity model (models 2, 3 and 4 in the results) looks as follows:

$$Log(Emigration)_{Spd,t} = \beta_0 + \beta_1 \text{Log}(Relative\ GDP\ pc)_{Spd,,t-1} + \beta_2 Log(Distance)_{Spd,t-1} + \beta_3 Log(Population\ Spain_{t-1} * Population\ destination_{t-1}) + \beta_4 Relative\ inflation_{Spd,t-1} + \beta_5 Language + \beta_k Augmented_{k,Spd,t-1} + \beta_k Welfare_{k,Spd,t-1} + \beta_k Relations_{k,Spd,t-1} + \beta_{dum} Region\ dummies + \delta_{Spd} + \varepsilon_{Spd,t} \quad (3.4)$$

where $Augmented_{k,Spd,t-1}$ refers to the variables added in model 2 (Relative Unemployment, Relative Youth Unemployment, Relative Inequality (Gini) and the Square of Relative Inequality). $Welfare_{k,Spd,t-1}$ refers to the variables added in model 3, that are Relative Educational Attainment, Relative Life Expectancy, Relative Expenditure on Education, Relative Expenditure on Health and Relative Taxes. $Relations_{k,Spd,t-1}$

refers to Log(Migrant Stock), Relative FDI and Trade. Finally, the dummies for the regions refer to the European Union, rest of Europe, Africa and Latin America.

The inclusion of unemployment in model 2 is in line with the theory of international migration, more specifically the extension of the net return formula by Sjaastad (1962), which assumes that the probability of the migrant to be employed affects their decision. Therefore, the probability of being unemployed affects also, both in the origin and the destination country. In the case of Spain, this has been proved empirically by Izquierdo et al. (2015). As for the inclusion of inequality, it has not been done for Spain yet, therefore I aim to expand the literature on this topic by following both the theory and the empirics. The inclusion of a relative term of inequality (which is measured with the Gini index in the main models) is in line with the theory of Borjas (1987) and the empirical research of Mayda (2010). Both of them relate inequality in the origin to inequality in the destination, and find that when the first one has high levels (that is, inequality in Spain in high), if the relative term increases, emigration decreases. The square of inequality is included due to the non-linear relation found by Mayda (2010) and Clark et al. (2007).

The welfare variables from model 3 are included following Bodvarsson et al. (2014) and Enache and Pânzaru (2012), although not all the variables the last ones include are significant. Education is part of this model as Stark and Taylor (1991) theorize that higher education in a country decreases its emigration rates. As for the variables in model 4, they all measure the relations between origin and destination countries in different ways. Relative FDI measures the openness of the countries, as the more open a country is, the easier it will be to migrate to and from it, while the variable Trade measures directly the extent of the economic relations between Spain and the destination countries. Migrant stocks, on the other hand, affect directly to the individuals who migrate, as addressed by the theory of migration networks (Moreno Torres and López Casasnovas, 2006; and Bodvarsson et al., 2014). Finally, the regional dummies are included in an effort to distinguish those regions that attract more migrants from Spain, both in general and when dividing the dataset in nationals and non-nationals.

3.2.3. Multilateral resistance and expected signs

An important issue in panel data migration models in multilateral resistance. This problem arises because the attractiveness of other destinations has to be measured when

estimating migration flows from Spain to one destination country: individuals not only take into account those two countries, but also all the other possibilities. This problem is also addressed in international trade models, as Bacchetta et al. (2012) and Anderson and Van Wincoop (2003) discuss.

While Mayda proxies Multilateral Resistance to Migration using a weighted average of distance and GDP per worker of all the other possible destinations, Bertoli and Fernández-Huertas Moraga (2011) and Izquierdo et al. (2015) rely on a Common Correlated Effects estimator developed by Pesaran. However, Bertoli and Fernández-Huertas Moraga (2011) base their study on a Random Utility Maximization model, as explain in Section 2.2. This model is not being applied here. They also do not consider Mayda's approach general enough, however, they always use more than one origin country when trying to address Multilateral Resistance. As what it is in interest here is the attractiveness of the 60 destination countries for Spanish emigrants, a weighted average as follows (which can also be seen in Bacchetta et al., 2012, for international trade) will be used for each year:

$$MR_i = \sum_d dist_{Spd} * GDP\ share_d \quad (3.5)$$

Where "i" is the destination country of interest and d refers to the other possible destinations. The higher this term is, the higher emigration flows are. It will be added in models 2, 3 and 4.

The expected signs of the variables can be seen below in Table 3.5, and at the end of Section 4, once Results are discussed, a new table with the realised signs can be found.

Table 3.5 - Expected signs of variables

Model	Hypothesis	Variable	Expected Sign
(1) Benchmark	H1	Log(Relative GDP pc)	-
		Log(Distance)	-
	H2	Log(Population Spain * Population destination)	+
		Relative Inflation	-
		Language	+
(2) Augmented	H3	Relative Unemployment	+
		Relative Youth Unemployment	+
		Relative Inequality (Gini or S80/S20)	-
		Square of Relative Inequality (Gini or S80/S20)	+
	Multilateral Resistance	Multilateral Resistance	+
(3) Welfare	H4	Relative Educational Attainment	+
		Relative Life Expectancy	-
	H5	Relative Expenditure on Education	-
		Relative Expenditure on Health	-
		Relative Taxes	-
(4) Relations	H6	Log(Migrant stock)	+
		Relative FDI	+
		Trade	+

4. Results

This section presents the results of the estimated models in two different blocks: in Section 4.1 I will explain the results when using the variable Log(Emigration), that is, the whole emigration flows from Spain to the destination countries, as well as an extra robustness check for relative deprivation. Then, in Section 4.2 I will explain how the results change when splitting the database of emigration flows into nationals and non-nationals. All the models included in Sections 4.1 and 4.2 are only a condensed version of the results. The complete results, with all the variables that have been left out in the analysis, can be seen in Appendix D.

Before explaining the final results, it is important to look at the correlation matrix that can be seen in Appendix B. The most important information to be taken from the correlation matrix is the high correlation between some independent variables, such as distance and language, GDP per capita and life expectancy and health expenditure, or population and trade and migrant stock. This leads to multicollinearity issues that may bias the estimation, especially because distance, GDP per capita and population are key variables for the gravity model and cannot be dropped.

Other problems that have been dealt with are endogeneity and reverse causality, which have already been addressed in Section 3.2 and dealt with by including lagged variables. As for the necessity of including country fixed effects to account for the unobserved characteristics of each country, the Hausman test gives different results depending on the dependent variable used, as will be explained later. Furthermore, the results of the heteroscedasticity and serial correlation tests can be found in Appendix C.

As for the estimations, all of them, either with country-FE or RE, were done using Feasible Generalized Least Squares, which allow to correct for heteroskedasticy (in all cases) and autocorrelation AR1 (in the cases of emigration and emigration of non-nationals). A Driscoll and Kraay (1998) robust estimator, as well as OLS Panel-Corrected Standard Errors were considered, but resulted in worse outcomes. None of the estimations report the R-squared because it is not bounded between 0 and 1 and therefore would not offer a clear explanation, nor log-likelihood as it is only shown in Stata when using an iterated GLS estimator instead of two and three-steps estimators (the latter in the case of autocorrelation). AIC or BIC estimators are also not reported in Stata.

4.1. General results

The results of the estimations with Log(Emigration) as dependent variable can be seen in Table 4.2. They consist of four equations that align with the hypothesis from Section 3.1: the benchmark model, the augmented model, welfare and relations. Not all variables are included here as many do not offer conclusive results or present correlation issues with other independent variables. It is the case of the Relative Youth Unemployment (although when included it is significant and has the expected sign), Language and especially welfare variables such as Relative Taxes (which when included shifts the sign of distance to positive), Relative Educational Attainment and Relative Life Expectancy.

Due to the discrepancies in the registrations of emigration flows of non-nationals that are part of total emigration flows, as was explained previously, a subset of the data from 2006 has been used to run the same regressions as a robustness check of the estimations for the complete dataset. It can be observed that most of the coefficients are very similar in both cases, although smaller for the subset. All the models are estimated with FGLS and country fixed effects, as suggested by the Hausman test that can be seen in Table 4.1.

Table 4.1 - Hausman test with Log(Emigration) as dependent variable

Test Summary	x^2 **Statistic**	***p*-value**
	377.47	0.0000***

Notes: (i) Ho: difference in coefficients not systematic.

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Looking at the results by blocks, column (1) shows the benchmark model without the variable Language. Log(Relative GDP per capita) has a negative sign and it is significant, which shows that when GDP per capita in Spain increases, emigration flows from this country decrease. This is in line with neoclassical theory (Moreno Torres and López Casasnovas, 2006) and the empirical findings of all the literature, like Son and Noja (2012), Mayda (2010) or Carlos (2002). Also negative is the coefficient of Log(Distance), while the interaction term of populations is positive, although both are significant at a 1% level. All of this is in line with the theory and empirics of the gravity model, like the results from Lewer and Van den Berg (2008), Ullah (2012) and Mayda (2010). However, Relative Inflation is not significant.

Continuing with column (2), it refers to the augmented model, that is, the hypotheses of unemployment and relative deprivation, that are added to the benchmark model. While the negative sign of Relative Unemployment is in line with the literature for Spain such as Izquierdo et al. (2015), that is, that higher relative unemployment in Spain increases emigration flows, it is only significant for the subset of data. This negative sign also follows the theory of international migration from Sjaastad (1962), Bodvarsson et al. (2014) and the empirical research of Beine et al. (2013).

As for relative deprivation, both Relative Gini and its square have the expected signs (negative and positive, respectively), which differ from usual positive results in the literature as Spain already has high relative inequality. However, it is in line with the theory of Borjas (1987) and Stark (1984) and the results of Mayda (2010), Clark et al. (2007) and Son and Noja (2012). Finally, the coefficient for Multilateral Resistance to Migration is positive and significant, although the sign is different to Mayda's (2010), as the measurement is different.

Moving on to the welfare variables in column (3), I find a few problems. First, as I indicated earlier some of them had to be taken out, and those who are included here have high correlation with Log(Relative GDP per capita), which could be the reason why although the signs are usually correct, only Relative Expenditure on Education is significant, as well as positive. Enache and Pânzaru (2012) also find that most welfare variables are not significant. Furthermore, the constant disappears and many observations drop out, although this last problem also seems to happen to Mayda (2010). Relative Inflation turns significant in this model and has a negative sign, which is line with the findings by Son and Noja (2012), who argue that higher inflation in the origin country reduces migrant's capacity to move and live abroad, as it increases migration costs.

However, when including the variables for Relations in column (4), Relative Expenditure on Education turns insignificant. While Relative FDI is positive and significant, meaning that the more open a country is, the more migration flows are registered, Log(Migrant Stock) is negative, which does not fit the theory or the empirical literature by Clark et al. (2007) and Rodríguez-Fariñas et al. (2016). I also expected Trade to be positive, however it is negative. This would be the right solution according to the theory by Mundell (1957), who finds that goods and factors (in this case it would be labour) movements are substitutes.

Table 4.2- Output for Log(Emigration) as dependent variable

VARIABLES	(1) Benchmark (H. 1 & 2)		(2) Augmented (H. 3)		(3) Welfare (H. 4 &5)		(4) Relations (H. 6)	
	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006
Log(Relative GDP pc)	-1.863***	-1.079***	-0.970***	-0.348	-1.183***	-0.720***	-1.260***	-0.995***
	(-10.18)	(-6.031)	(-3.835)	(-1.391)	(-4.843)	(-2.717)	(-5.027)	(-4.649)
Log(Distance)	-6.112***	-4.407***	-5.540***	-3.782***	-12.31***	-10.81***	-11.84***	-10.02***
	(-12.97)	(-7.960)	(-6.091)	(-4.111)	(-13.09)	(-10.36)	(-12.76)	(-11.61)
Log(Population Spain*Population destination)	5.235***	4.475***	5.062***	4.224***	6.288***	5.406***	6.168***	5.202***
	(18.91)	(13.67)	(11.01)	(9.206)	(14.08)	(10.32)	(14.06)	(12.35)
Relative Inflation	-0.00127	-0.000547	-0.000588	0.000139	-0.00124	-0.00192**	-0.00193**	-0.00121**
	(-1.466)	(-0.664)	(-0.577)	(0.163)	(-1.365)	(-2.088)	(-2.056)	(-2.083)
Relative Unemployment			0.00219	0.0489**	-0.0205	0.0452**	-0.00975	0.0211
			(0.104)	(2.513)	(-1.002)	(2.112)	(-0.485)	(1.210)
Relative Inequality (Gini)			-2.556**	-1.514	-5.111***	-4.054***	-5.056***	-4.274***
			(-2.175)	(-1.484)	(-4.167)	(-3.230)	(-4.054)	(-4.085)
Square of Relative Inequality (Gini)			1.116**	0.695	2.143***	1.686***	2.216***	1.856***
			(2.132)	(1.476)	(4.023)	(3.015)	(4.098)	(3.895)
Multilateral Resistance			0.0134**	0.0103*	0.0125*	0.0110*	0.00150	0.00928
			(2.092)	(1.854)	(1.885)	(1.802)	(0.224)	(1.537)
Relative Expenditure on Education					-0.214*	-0.222	-0.297***	-0.285**
					(-1.898)	(-1.507)	(-2.660)	(-2.083)
Relative Expenditure on Health					-0.0211	0.103**	-0.0602	0.0566
					(-0.515)	(2.335)	(-1.352)	(1.293)
Log(Migrant Stock)							0.0180	-0.0990**
							(0.206)	(-2.051)
Relative FDI							0.00876	0.00595*
							(1.530)	(1.723)
Trade							-0.0956***	-0.0991***
							(-4.405)	(-6.779)
Constant	-45.21***	-45.56***	-48.49***	-47.74***	-	-	-	-
	(-16.34)	(-16.07)	(-10.88)	(-11.24)				
Observations	580	455	398	359	302	269	302	269
Country FE	YES	YES	YES	YES	YES	YES	YES	YES

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) are lagged; (iv) all models are estimated with country FE and FGLS.*

I have included dummies for regions (not shown in the tables) which indicate that, for both datasets, migration flows to Europe and the EU and smaller compared to those to Latin America, especially in the second case (the coefficients vary from -6 to -41).

Finally, in Appendix D a table (D.6) with a robustness check for relative deprivation using the index S80/S20 is performed, showing very similar results in all columns and the dummies.

4.2. Extended specifications

Due to the situation of migration in Spain during the last two decades, it is interesting to differentiate between nationalities, as their destinations when emigrating vary greatly (see Section 3.1). Therefore, I decided to split the dependent variable into two subsets, one for Log(Emigration of nationals) (Table 4.5) and another for Log(Emigration of non-nationals) (Table 4.6). While in the second case, the estimation was the same as for total emigration flows (FGLS with country fixed effects and correcting for heteroskedasticy and serial correlation), in the case of nationals the Hausman test that can be seen in Table 4.3 shows that random effects are preferred and that there is no serial correlation. Therefore, as for non-nationals fixed effects are necessary (see Table 4.4), the estimation methods differ for each subset of the data. Again, as robustness check I decided to perform the same analysis from 2002 and 2006, to eliminate possible biases due to registration discrepancies.

Table 4.3 - Hausman test with Log(Emigration of nationals) as dependent variable

Test Summary	x^2 **Statistic**	***p*-value**
	2.64	0.4507***

Notes: (i) Ho: difference in coefficients not systematic.

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Table 4.4 - Hausman test with Log(Emigration of non-nationals) as dependent variable

Test Summary	x^2 **Statistic**	***p*-value**
	16.66	0.0008***

Notes: (i) Ho: difference in coefficients not systematic.

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

As can be seen, the results in both cases are not as conclusive as earlier. In fact, for nationals the benchmark model is never significant. However, excluding the Trade variable, all signs are correct and coincide with the estimation using total emigration flows. Furthermore, when adding dummies for regions, they are also not significant, but positive in the case of the EU, which would indicate that nationals emigrate to this region more with respect to Latin America.

Regarding the estimation using emigration of non-nationals, which does use fixed effects, the results are better, although not as good as with total emigration flows. The benchmark models are mostly significant and with the right signs, although the augmented, welfare and relations variables again fail to give any significant results. As for the regional dummies for Europe and the EU, they are both negative and significant.

A possible explanation for the low quality of the welfare and relations models is the fact that there seem to be correlation issues, as many observations drop out when running the regressions. However, many of the high correlations that can be seen in the matrix in Appendix B involve GDP per capita, distance and population, which are the baseline of the gravity model, making it harder to correct the problem.

Table 4.5- Output for Log(Emigration of nationals) as dependent variable

VARIABLES	(1) Benchmark (H. 1 & 2)		(2) Augmented (H. 3)		(3) Welfare (H. 4&5)		(4) Relations (H. 6)	
	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006
Log(Relative GDP pc)	-0.140	-0.0268	-0.267	-0.256	-0.596	-0.608	-0.526	-0.386
	(-0.589)	(-0.100)	(-1.054)	(-0.931)	(-1.287)	(-1.309)	(-0.981)	(-0.672)
Log(Distance)	0.359	0.183	-0.0962	-0.0916	-0.426	-0.207	-1.090	-1.079
	(0.975)	(0.460)	(-0.174)	(-0.160)	(-0.748)	(-0.312)	(-1.523)	(-1.298)
Log(Population Spain*Population destination)	0.139	0.0741	0.0963	0.106	0.0552	0.0859	0.171	0.166
	(0.983)	(0.459)	(0.614)	(0.631)	(0.311)	(0.460)	(0.735)	(0.678)
Relative Inflation	-0.0105**	-0.00625	-0.0122**	-0.00879*	-0.0144**	-0.0110*	-0.0145**	-0.0115*
	(-2.312)	(-1.311)	(-2.201)	(-1.671)	(-2.221)	(-1.792)	(-2.157)	(-1.758)
Relative Unemployment			0.0806	0.0764	0.125	0.0907	0.131	0.104
			(0.596)	(0.615)	(0.772)	(0.602)	(0.799)	(0.695)
Relative Inequality (Gini)			-14.94**	-14.96**	-20.16***	-17.05**	-18.15**	-14.03**
			(-2.131)	(-2.143)	(-2.887)	(-2.353)	(-2.575)	(-2.087)
Square of Relative Inequality (Gini)			6.301*	6.380**	8.950***	7.586**	8.070**	6.382**
			(1.901)	(1.979)	(2.686)	(2.236)	(2.422)	(2.001)
Multilateral Resistance			0.0116	-0.00327	-0.0347	-0.0366	-0.0464	-0.0573
			(0.304)	(-0.0920)	(-0.795)	(-0.853)	(-1.085)	(-1.328)
Relative Expenditure on Education					1.264**	0.683	1.208*	0.603
					(1.986)	(1.072)	(1.801)	(0.921)
Relative Expenditure on Health					0.257	0.261	0.275	0.269
					(1.324)	(1.142)	(1.311)	(1.078)
Log(Migrant Stock)							0.157	0.243
							(0.802)	(1.169)
Relative FDI							0.0419	0.0691
							(0.888)	(1.328)
Trade							-0.256***	-0.315***
							(-2.582)	(-3.017)
Constant	-1.653	0.951	9.479	11.20	19.84*	16.68	22.47*	21.55*
	(-0.417)	(0.207)	(0.917)	(1.136)	(1.821)	(1.447)	(1.946)	(1.662)
Observations	563	445	386	352	293	263	293	263
Country FE	NO	NO	NO	NO	NO	NO	NO	NO

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) are lagged; (iv) all models are estimated with RE as indicated by Hausman test.*

Table 4.6- Output for Log(Emigration of non-nationals) as dependent variable

VARIABLES	(1) Benchmark (H. 1 & 2)		(2) Augmented (H. 3)		(3) Welfare (H. 4&5)		(4) Relations (H. 6)	
	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006
Log(Relative GDP pc)	-1.011**	-0.341	-0.891*	-0.323	-1.197	0.143	-0.770	0.342
	(-2.298)	(-1.295)	(-1.900)	(-0.847)	(-1.559)	(0.194)	(-0.975)	(0.450)
Log(Distance)	-3.788***	-1.032	-4.691**	-5.558***	-8.928***	-8.339***	-8.938***	-7.256**
	(-3.394)	(-1.148)	(-2.485)	(-3.121)	(-2.994)	(-2.607)	(-2.902)	(-2.217)
Log(Population Spain*Population destination)	3.851***	2.499***	4.319***	4.874***	4.576***	4.043***	4.489***	3.469**
	(5.825)	(4.806)	(4.150)	(5.154)	(3.150)	(2.605)	(2.977)	(2.149)
Relative Inflation	0.00270	0.00122	0.00158	0.000188	-0.00136	-0.000434	-0.00138	0.000700
	(1.420)	(0.876)	(0.729)	(0.108)	(-0.472)	(-0.170)	(-0.420)	(0.237)
Relative Unemployment			-0.0543	-0.0257	0.00808	0.0583	0.0251	0.0804
			(-1.136)	(-0.619)	(0.136)	(1.085)	(0.410)	(1.600)
Relative Inequality (Gini)			-2.676	-2.728	-4.554	-3.033	-5.049	-3.082
			(-1.099)	(-1.227)	(-1.167)	(-0.812)	(-1.246)	(-0.786)
Square of Relative Inequality (Gini)			1.025	0.948	1.626	0.982	1.816	0.934
			(0.939)	(0.927)	(0.971)	(0.606)	(1.065)	(0.564)
Multilateral Resistance			0.0120	0.00703	0.00704	0.0102	0.00718	0.0158
			(0.893)	(0.753)	(0.398)	(0.691)	(0.357)	(0.874)
Relative Expenditure on Education					0.257	-0.191	0.181	-0.199
					(0.521)	(-0.389)	(0.356)	(-0.398)
Relative Expenditure on Health					0.0792	0.121	0.111	0.161
					(0.486)	(0.696)	(0.665)	(0.915)
Log(Migrant Stock)							0.0955	-0.0345
							(0.576)	(-0.234)
Relative FDI							0.00158	0.00845
							(0.142)	(0.782)
Trade							0.226	0.248
							(1.049)	(1.051)
Constant	-39.36***	-36.38***	-41.14***	-45.29***	-	-	-	-
	(-6.735)	(-8.965)	(-5.657)	(-7.461)				
Observations	578	454	397	359	299	265	299	265
Country FE	YES	YES	YES	YES	YES	YES	YES	YES

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) are lagged; (iv) all models are estimated with country FE and FGLS.*

As can be seen in Table 4.7, all the variables except for Trade have the same expected and realised signs, and the result for that variable has an explanation in the theory of Mundell (1957).

Table 4.7 – Realised signs versus expected signs

Model	Hypothesis	Variable	Realised Sign	Expected Sign
(5) Benchmark	H1	Log(Relative GDP pc)	-	-
		Log(Distance)	-	-
	H2	Log(Population Spain * Population destination)	+	+
		Relative Inflation	-	-
		Language	+	+
(6) Augmented	H3	Relative Unemployment	+	+
		Relative Youth Unemployment	+	+
		Relative Inequality (Gini or S80/S20)	-	-
		Square of Relative Inequality (Gini or S80/S20)	+	+
	Multilateral Resistance	Multilateral Resistance	+	+
(7) Welfare	H4	Relative Educational Attainment	+/-	+
		Relative Life Expectancy	-	-
	H5	Relative Expenditure on Education	-	-
		Relative Expenditure on Health	-/+	-
		Relative Taxes	-	-
(8) Relations	H6	Log(Migrant stock)	+	+
		Relative FDI	+	+
		Trade	-	+

Notes: (i) The signs of the variables Language, Relative Youth Unemployment, Relative Inequality using S80/S20, Relative Educational Attainment, Relative Life Expectancy and Relative Taxes can be found in Appendix D.

4.3. Policy implications

The high levels of emigration registered in Spain can have consequences for the economy that have to be considered when dealing with this issue.

The first issue is the relation between emigration and unemployment. I obtain a positive relation between unemployment in Spain and emigration, however, high levels of emigration can reduce levels of unemployment, although it is not clear to what extent as

many of the studies assume that emigrants are unemployed indivivuals and do not consider regional differences, which, as seen in Section 3.1, can be quite important in the case of Spain. It is seems clearer, however, that emigration increases wages in the country of origin (Asch adn Reichmann, 1994).

Other consequences of emigration from Spain are related to the age group that decides to migrate. Several studies agree on the serious consequences of young emigration that will not return as long as the economic situation does not improve, which could derive in a "lost generation" problem in the future (Navarrete Moreno et al., 2014). Furthermore, when these emigrants are highly qualified people, the government's investment in their education does not generate any benefits.

The overall negative influence of inequality on emigration is due to the fact that those that are already in the upper tail of the distribution do not have incentives to migrate to countries with lower inequality, as they would not improve their position, and they would have less incentives to migrate if inequality increases more (Mayda, 2010). It should be pointed out that the comparison with Europe and Latin America is different. Latin America's levels of inequality are higher than those in Spain, which would suggest a positive selection of migrants (although a big part of these migrants are actaually returning to their countries of origin), while Europe's levels of inequality differ by region, therefore encounerting both positive and negative selection. Therefore, when analysing this issue it is important to take these differences into account.

Finally, it is important to point out that the policy implications of these results are not clear, as the positive effects of emigration have not been studied in depth for developed countries like Spain (Izquierdo et al., 2015).

5. Conclusion

The thesis has covered the topic of international migration, focusing on flows from Spain to 60 destination countries from 2002 to 2013. This is a topic of vital importance, as there are currently over 2 million Spanish citizens living abroad, which is an increase of 56% from 2002 to 2013. During the last century, Spain's migration history has gone from phases of high emigration rates to net entries of immigrants, and, with the last crisis, back to emigration. Since this crisis hit Spain in 2008, GDP pc growth plummeted, inequality increased and unemployment reached 26.9% in 2013, while some sources have registered up to 400,000 emigrants per year, though a great portion of this emigration is composed former immigrants that return to their countries) Although non-nationals emigration outweighs nationals' emigration, the second has caught more attention, generating alarm of a possible "brain drain" and "lost generation".

These new Spanish emigration trends have not been researched broadly, due to them being so recent. Although there are some studies, such as the ones by as Izquierdo et al. (2015), Navarrete Moreno et al. (2014) and Rodríguez-Fariñas et al. (2016), none of them include inequality as an important determinant of emigration. With this paper I aimed to expand the literature by adding a longer timespan (going back to 2002) and focusing on European and American countries. The time is limited due to the lack of data on emigration flows with a known destination country. As inequality and welfare as variables that influence emigration have not been studied in depth for Spain, I included them in this paper.

The results of the gravity model are not all fully conclusive. I find that the benchmark model, which is based on the effects of GDP per capita, distance and population, is always significant and has the correct signs: higher relative GDP per capita in Spain decreases emigration flows, same as distance, while bigger populations increase those flows. However, the same does not happen with all the variables of the augmented model. I find that higher relative unemployment in Spain increases emigration flows, and higher relative inequality decreases them. This last variable has a different sign than in the literature as Spain already has high inequality levels. However, both the welfare and relations blocks do not have the expected significant results, and when I split the database to distinguish between nationals and non-nationals, the models do not offer the expected significant results either, suggesting that there should be more research.

There are several ways to expand and improve this analysis. The first one is by using data that does not suffer from the discrepancies that *Estadística de Variaciones Residenciales* has. This is difficult because although *Estadística de Migraciones* is a better dataset that does not suffer from changes in deregistration laws, it covers only from 2008. Furthermore, as regions in Spain have very different characteristics, it would be better to study emigration flows from different regions to destination countries. Again, currently this is not possible due to data restrictions, but it would give a better overview of international migration in Spain. Furthermore, here I only distinguish between nationals and non-nationals, however, the same analysis can be done for people born in Spain or outside, which may result in different estimates as many immigrants are able to obtain the Spanish nationality and then return to their countries of origin: return migration might be playing a big role in the overall emigration flows.

Other variables could also be added to the analysis. As has been seen, welfare and relations do not perform well in this model, however other measures could be used. For example, studying the influence of pension systems on migration or migration policy, including similarities and differences of law systems, would improve the analysis. Furthermore, as Spain was hit by a housing bubble that worsened the crisis in 2008, real estate prices could also be included.

Finally, research of international migration should not be limited to its causes, but it should also be extended to the consequences. How much can "brain drain" hurt the economy? Izquierdo et al. (2015) highlight the importance of this, which has been in Spanish newspapers since the phenomenon started. But, can emigration be beneficial for Spain? Opportunities abroad may be an incentive to improve education, emigrants send remittances back home (and they may return themselves in the future), and scientific networks can be formed.

Although there is much research left to be done, this thesis contributes to the literature regarding Spanish migration, as it adds the importance of relative deprivation to the previously studied unemployment and income differences.

6. Bibliography

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics*, *3*(1), 133-160.

Anderson, J. E., and Van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *The American Economic Review*, *93*(1), 170-192.

Asch, B. J., & Reichmann, C. (Eds.). (1994). *Emigration and its effects on the sending country*. Rand.

Bacchetta, M., Beverelli, C., Cadot, O., Fugazza, M., Grether, J. M., Helble, M., Nicita, A., and Piermartini, R. (2012). *Chapter 3: Analyzing Bilateral Trade using the Gravity Equation. A Practical Guide to Trade Policy Analysis*. United Nations and World Trade Organization.

Baum, C. F. (2001). Residual diagnostics for cross-section time series regression models. *The Stata Journal*, 1(1), 101-104

Beine, M. A., Bricongne, J. C., and Bourgeon, P. (2013). Aggregate fluctuations and international migration. *CESifo Working Paper No. 4379.*

Belot, M. V., and Hatton, T. J. (2012). Immigrant Selection in the OECD. *The Scandinavian Journal of Economics*, *114*(4), 1105-1128.

Bertoli, S., Brücker, H., and Fernández-Huertas Moraga, J. (2013). *The European Crisis and Migration to Germany: Expectations and the Diversion of Migration Flows* (No. 7170). Institute for the Study of Labor (IZA).

Bertoli, S., and Fernández-Huertas Moraga, J. (2011). Multilateral Resistance to Migration. *Documento de Trabajo*, 04.

Blau, F. D., and Kahn, L. M. (2014). Immigration and the Distribution of Incomes. *Handbook of the Economics of International Migration: The Impact*, 793.

Bodvarsson, O. B., Simpson, N. B., and Sparber, C. (2014). Migration Theory. *Handbook of the Economics of International Migration, 1A: The Immigrants*, 1.

Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *The American Economic Review*, 531-553.

Carlos, R. D. (2002). On the Determinants of International Migration in the Philippines: An Empirical Analysis. *International Migration Review*, *36*(1), 81-102.

Chiswick, B. R., and Miller, P. W. (2014). International migration and the economics of language. *Handbook of the Economics of International Migration. Amsterdam: Elsevier*, 211-270.

Clark, X., Hatton, T. J., and Williamson, J. G. (2007). Explaining US immigration, 1971-1998. *The Review of Economics and Statistics*, *89*(2), 359-373.

Czaika, M., and de Haas, H. (2012). The role of internal and international relative deprivation in global migration. *Oxford Development Studies*, *40*(4), 423-442.

Domingo i Valls, A., and Sabater Coll, A. (2012). Crisis económica y emigración: la perspectiva demográfica. *Anuario de la Inmigración en España*, (2012), 59-88.

Domingo i Valls, A., Sabater Coll, A, and Ortega Rivera, E. (2014). ¿Migración neohispánica? El impacto de la crisis económica en la emigración española. *Empiria. Revista de metodología de ciencias sociales*, (29), 39-66.

Driscoll, J. C., and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, *80*(4), 549-560.

Enache, C., and Pânzaru, C. (2012). Romanian Migration Flows In European Countries: Does Social Security Matter? *Annales Universitatis Apulensis: Series Oeconomica*, *14*(2), 485.

Felbermayr, G. J., Grossmann, V., and Kohler, W. (2014). Migration, international trade and capital formation: Cause or effect? *Handbook of the Economics of International Migration. Amsterdam: Elsevier*, 913-1025.

Fernández, D., and Vitores, D. F. (2013). El Español: una lengua viva. *Instituto Cervantes.*

Fernández-Tabales, A., and Cruz, E. (2013). Análisis territorial del crecimiento y la crisis del sector de la construcción en España y la Comunidad Autónoma de Andalucía. *EURE (Santiago)*, *39*(116), 5-37.

Gómez, S., Contreras, I., and Gracia, L. (2008). Las reformas laborales en España y su impacto real en el mercado de trabajo en el período 1985-2008. *Documento de investigación DI-764, Cátedra Seat de Relaciones Laborales, IESE Business School*.

Greenwood, M. J. (1985). Human migration: Theory, models, and empirical studies. *Journal of regional Science*, *25*(4), 521-544.

Hanson, G. H. (2010). International Migration and the Developing World. *Handbook of Development Economics, Volume 5.*

Izquierdo, M., Jimeno, J. F., and Lacuesta, A. (2014). La emigración de españoles durante la Gran Recesión (2008-2013). *Cuadernos económicos de ICE,* (87), 223-240.

Izquierdo, M., Jimeno, J. F., and Lacuesta, A. (2015). Spain: From Immigration to Emigration? *Documentos de Trabajo. N.º 1503. Banco de España.*

Jakobsson, A. (2006). Trade Openness and Income Inequality. Lund University.

Jauer, J., Liebig, T., Martin, J., and Puhani, P. (2014). *Migration as an Adjustment Mechanism in the Crisis? A Comparison of Europe and the United States* (No. 155). OECD Publishing.

Jennissen, R.P.W. (2004). *Macro-economic Determinants of International Migration in Europe.* Rozenberg Publishers.

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

Liebig, T., and Sousa-Poza, A. (2004). Migration, Self-Selection and Income Inequality: An International Analysis. *Kyklos*, *57*(1), 125-146.

López-Bazo, E., and Motellón, E. (2013). The regional distribution of unemployment: What do micro-data tell us? *Papers in Regional Science*,*92*(2), 383-405.

Márquez Arboleda, L., Rochina Barrachina, M. E., and Antuñano Maruri, I. (2004). Un modelo de gravedad ampliado para la inmigración internacional en España. *Revista Valenciana de Economía y Hacienda*, *3*(12), 149-169.

Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, *23*(4), 1249-1274.

Merkus, E. (2015). Is the Grass Really Fairer on the Other Side? Migration and Inequality in Canada. *Erasmus University Rotterdam*.

Mihi-Ramírez, A., Rudžionis, A., and Kumpikaitė, V. (2014). European economic migration flow, earnings and unemployment in decade of 2000. *Procedia-Social and Behavioral Sciences*, *110*, 122-129.

Mundell, R. A. (1957). International trade and factor mobility. *The American Economic Review*, *47*(3), 321-335.

Navarrete Moreno, L., Cuenca García, C., Díaz Catalán, C., Díaz Chorne, L., Arcadio Flores-Vidal, P., Gentile, A., and Zúñiga Contreras, R. (2014). *La emigración de los jóvenes españoles en el contexto de la crisis: Análisis y datos de un fenómeno difícil de cuantificar*. INJUVE-Observatorio de la Juventud en España.

Ramos, C., Moreno, B., and Alvargonzález, M. (2015). Análisis de la evolución de la desigualdad de renta en las Comunidades Autónomas (2003-2012). *Reunión de Estudios Regionales.*

Ramos, R., and Royuela, V. (2016). *Graduate migration in Spain: the impact of the great recession on a low mobility country* (No. 201608). University of Barcelona, Research Institute of Applied Economics.

Ramos, R., and Suriñach, J. (2013). A gravity model of migration between ENC and EU. *IREA–Working Papers, 2013, IR13/017*.

Ravenstein, E. G. (1889). The laws of migration. Journal of the royal statistical society, 52(2), 241-305.

Rodríguez-Fariñas, M. J., Romero-Valiente, J. M., and Hidalgo-Capitán, A. L. (2016). Los exiliados económicos. La nueva emigración española a México (2008-2014). *Scripta Nova. Revista Electrónica de Geografía y Ciencias Sociales*, *20*.

Sjaastad, L. A. (1962). The Cost and Retums to Human Migration. *Journal of Political Economy*, *70* (5, Part 2), 80-93.

Son, L., and Noja, G. G. (2012). A macroeconometric panel data analysis of the shaping factors of labour emigration within the European Union. *Theoretical and Applied Economics*, *18*(11 (576)), 15-30.

Stark, O. (1984). Rural-to-urban migration in LDCs: a relative deprivation approach. *Economic Development and Cultural Change*, *32*(3), 475-486.

Stark, O., and Taylor, J. E. (1991). Migration incentives, migration types: The role of relative deprivation. *The economic journal*, *101*(408), 1163-1178.

Sulaimanova, B., and Bostan, A. (2014). International Migration: A Panel Data Analysis of the Determinants of Emigration from Tajikistan and Kyrgyzstan. *Eurasian Journal of Business and Economics*, *7*(13), 1-9.

Tilly, C. (2011). The impact of the economic crisis on international migration: a review. *Work, Employment and Society*, *25*(4), 675-692.

Tinbergen, J., 1962. *Shaping the World Economy*. Twentieth Century Fund, New York, NY.

Moreno Torres, I., and López Casasnovas, G. (2006). Los determinantes de la inmigración internacional en España. *Investigaciones Regionales*, (9), 25-48.

Ullah, M. S. (2012). Determinants of international labor migration from Bangladesh: a gravity model of panel data.

http://www.ritsumei.ac.jp/acd/re/ssrc/result/memoirs/kiyou25/25-06.pdf (Accessed: May 18th 2016)

UN-DESA, O. E. C. D. (2013). World migration in figures. *A Joint Contribution by UNx2010DESA and the OECD to the United Nations Highx2010Level Dialogue on Migration and Development*.

Williamson, J. G. (2014). World Migration in Historical Perspective: Four Big Issues. *Handbook of the Economics of International Migration, 1A: The Immigrants*, 89.

Zipf, G. K. (1946). The P_1P_2/D hypothesis: on the intercity movement of persons. *American Sociological Review*, *11*(6), 677-686.

Appendix A

Table A.1 - List of countries and regions

Country	Region
Austria	EU
Belgium	EU
Bulgaria	EU
Croatia	EU
Cyprus	EU
Czech Republic	EU
Denmark	EU
Estonia	EU
Finland	EU
France	EU
Germany	EU
Greece	EU
Hungary	EU
Ireland	EU
Italy	EU
Latvia	EU
Lithuania	EU
Luxembourg	EU
Malta	EU
Netherlands	EU
Poland	EU
Portugal	EU
Romania	EU
Slovakia	EU
Slovenia	EU
Sweden	EU
United Kingdom	EU

Country	Region
Albania	Europe
Andorra	Europe
Belarus	Europe
Bosnia	Europe
Iceland	Europe
Liechtenstein	Europe
Macedonia	Europe
Moldova	Europe
Norway	Europe
Russia	Europe
Switzerland	Europe
Ukraine	Europe
Algeria	Africa
Morocco	Africa
Argentina	America
Bolivia	America
Brazil	America
Chile	America
Colombia	America
Costa Rica	America
Cuba	America
Dominican Republic	America
Ecuador	America
El Salvador	America
Guatemala	America
Honduras	America
Mexico	America
Nicaragua	America
Panama	America
Paraguay	America
Peru	America
Uruguay	America
Venezuela	America

Table A.2 – List of variables, descriptions and sources (Part 1)

Type	Variable	Description	Data Source
Dep.	Log(Emigration)	Logarithm of total Emigration flows from Spain to known destination countries.	Instituto Nacional de Estadística: Estadística de Variaciones Residenciales (EVR). http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft20%2Fp307&file=inebase&L=0
	Log(Emigration of nationals)	Logarithm of Emigration flows of nationals from Spain to known destination countries.	
	Log(Emigration of non-nationals)	Logarithm of Emigration flows of non-nationals from Spain to known destination countries.	
H1	Log(Relative GDP pc)	Logarithm of ratio of GDP per capita in Spain over GDP per capita in destination country (constant US$).	World Bank World Development Indicators database. http://data.worldbank.org/data-catalog/world-development-indicators
	Log(Distance)	Logarithm of Distance (in km) between Madrid and capitals of destination countries.	Timeanddate. http://www.timeanddate.com/worldclock/distance.html
H2	Log(Population Spain * Population destination)	Logarithm of interaction term Population in Spain * Population in destination country (in thousands).	World Bank World Development Indicators database. http://data.worldbank.org/data-catalog/world-development-indicators
	Relative Inflation	Ratio of Inflation in Spain over Inflation in destination country (measured as GDP deflator, in %).	World Bank World Development Indicators database. http://data.worldbank.org/data-catalog/world-development-indicators
	Language	Dummy variable with value 1 for Spanish, 0 for the rest.	Instituto Cervantes. http://www.cervantes.es/imagenes/File/prensa/El%20espaol%20una%20lengua%20viva.pdf
H3	Relative Unemployment	Ratio of Unemployment in Spain over Unemployment in destination country (measured over share of labor force).	World Bank World Development Indicators database. http://data.worldbank.org/data-catalog/world-development-indicators
	Relative Youth Unemployment	Ratio of Youth Unemployment Spain over Youth Unemployment in destination country (measured as Share of youth not in education, employment, or training over total of youth population).	
	Relative Inequality (Gini or S80/S20)	Ratio of Gini or S80/S20 index in Spain over Gini or S80/S20 index in destination country (Gini ranges from 1 to 100).	
	Square of Relative Inequality (Gini or S80/S20)	Square of ratio of Gini or S80/S20 index in Spain over Gini or S80/S20 index in destination country (Gini ranges from 1 to 100).	

Table A.3 – List of variables, descriptions and sources (Part 2)

Type	Variable	Description	Data Source
H4	Relative Educational Attainment	Ratio of Educational attainment in Spain over Educational attainment destination country (measured as % of population over 25 years old with least post-secondary education completed).	UNESCO Unesco Institute of Statistics Database. http://data.uis.unesco.org/
	Relative Life Expectancy	Ratio of Life expectancy in Spain over Life expectancy in destination country.	World Bank World Development Indicators database. http://data.worldbank.org/data-catalog/world-development-indicators
H5	Relative Expenditure on Education	Ratio of Government expenditure on education in Spain over Government expenditure on education in destination country (measured as % of GDP).	World Bank World Development Indicators database. http://data.worldbank.org/data-catalog/world-development-indicators
	Relative Expenditure on Health	Ratio of Health expenditure per capita in Spain over Health expenditure per capita in country (measured in PPP, constant 2011 international $).	
	Relative Taxes	Ratio of Tax revenue in Spain over Tax revenue in destination country (measured as % of GDP).	
H6	Log(Migrant stock)	Logarithm of stock of Spanish emigrants in destination country.	Instituto Nacional de Estadística: Censo Electoral de los Residentes Ausentes. http://www.ine.es/ss/Satellite?c=Page&cid=1254735793323&pagename=CensoElectoral%2FINELayout&L=0
	Relative FDI	Ratio of FDI inflows in Spain over FDI inflows in destination country (nominal million $).	UNCTAD Database. http://unctadstat.unctad.org/EN/Index.html
	Trade	Ratio of Trade of destination country with Spain over World Trade of Spain (measured as Exports + Imports over GDP).	Data on Exports and Imports from Estadísticas del Comercio Exterior http://datacomex.comercio.es/ GDP from World Bank World Development Indicators database. http://data.worldbank.org/data-catalog/world-development-indicators
Dummies	Europe, Africa, EU, America	Dummy variables for regions.	
	Multilateral Resistance	$\sum_d dist_{Spd} * GDP\ share_d$	Data for Distance, see variable Log(Distance). Data for GDP, see variable Trade.

Notes: (i) subscript "Sp" indicates Spain; subscript "d" indicates destination country

Table A.4 – Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Log(Emigration)	645	5.733685	1.832219	0	10.27664
Log(Emigration of nationals)	627	3.778641	2.016898	0	6.876265
Log(Emigration of non-nationals)	643	3.838119	2.07988	0	6.882438
Log(Relative GDP pc)	657	1.035719	1.23442	-1.473296	3.630873
Log(Distance)	720	7.983556	0.8285592	6.202536	9.276128
Language	720	0.3166667	0.4654996	0	1
Relative Inflation	631	13.74174	9.244343	-5.657937	28.46914
Log(Population Spain*Population destination)	660	19.66608	1.644894	14.15801	22.97107
Relative Gini	426	0.9840706	0.23901	0.543775	1.470468
Relative S80/S20	426	1.06138	0.5588347	7.98E-10	2.572868
Relative Unemployment	638	2.193421	1.448977	0.2388889	10.64706
Relative Youth Unemployment	377	1.457688	0.8193352	0.2864407	5.229226

Figure A.1 - Map of Spain's regions (Comunidades autónomas)

Appendix B

Table B.1 - Correlation matrix: Part 1

	Log(Emigration)	Log(Emigration of nationals)	Log(Emigration of non-nationals)	Log(Relative GDP pc)	Log(Distance)	Language	Relative Inflation	Log(Population Spain*Population destination)	Relative Inequality (Gini)	Relative Inequality (S80/S20)	Relative Unemployment
Log(Emigration)	1										
Log(Emigration of nationals)	0.073	1									
Log(Emigration of non-nationals)	-0.0276	0.1851	1								
Log(Relative GDP pc)	-0.1496	-0.054	-0.1736	1							
Log(Distance)	-0.2983	0.0385	-0.1501	0.5157	1						
Language	0.1553	0.1655	-0.1649	0.5085	0.704	1					
Relative Inflation	0.0262	-0.0405	-0.0202	-0.0874	-0.0508	-0.0161	1				
Log(Population Spain*Population destination)	0.8218	0.0391	-0.1181	-0.0153	-0.053	0.3495	0.0586	1			
Relative Inequality (Gini)	-0.1469	0.0079	0.2202	-0.4412	-0.3037	-0.6717	-0.0283	-0.3024	1		
Relative Inequality (S80/S20)	0.2284	-0.0074	-0.1973	0.3681	0.377	0.6596	0.3851	0.3296	-0.5583	1	
Relative Unemployment	0.0771	0.1868	0.1508	-0.1056	0.1603	0.1185	-0.1015	-0.0844	0.308	0.0671	1
Relative Youth Unemployment	0.0863	0.2756	0.2964	-0.5019	-0.1782	-0.2484	0.0121	-0.0952	0.5122	-0.1821	0.6385
Relative Education Attainment	0.2781	0.2472	-0.1307	0.3087	-0.0501	0.4129	0.0381	0.4783	-0.3819	0.3341	-0.1651
Relative Life Expectancy	-0.2245	-0.0867	-0.1757	0.9248	0.5472	0.4281	-0.0967	-0.1527	-0.3716	0.2969	-0.0754
Relative Expenditure on Education	0.2149	0.1734	-0.1092	0.2777	0.2022	0.3896	-0.0108	0.3063	-0.1898	0.2197	0.1489
Relative Expenditure on Health	-0.1618	-0.0211	-0.088	0.9186	0.6179	0.5768	-0.1187	-0.0757	-0.4245	0.3524	0.0538
Relative Taxes	0.1619	-0.264	-0.1355	0.2901	0.0728	0.189	-0.0792	0.253	-0.306	0.0907	-0.22

Table B.2 - Correlation matrix: Part 2

	Log(Emigration)	Log(Emigration of nationals)	Log(Emigration of non-nationals)	Log(Relative GDP pc)	Log(Distance)	Language	Relative Inflation	Log(Population Spain* Population destination)	Relative Inequality (Gini)	Relative Inequality (S80/S20)	Relative Unemployment
Log(Migrant stock)	0.7657	0.0358	-0.0229	-0.4969	-0.2925	0.2042	0.0764	0.7444	-0.1497	0.2028	0.058
Relative FDI	-0.0205	-0.1274	-0.1631	-0.1136	-0.0527	-0.1272	-0.0668	0.0201	0.001	0.0181	-0.0909
Trade	0.569	-0.2409	-0.1252	-0.3452	-0.5033	-0.1809	0.0518	0.617	-0.076	-0.0459	-0.2325
Multilateral Resiatance	0.2216	-0.064	-0.0946	0.4323	0.0416	0.1403	-0.2596	0.1307	0.0263	0.2262	0.3295

	Relative Youth Unemployment	Relative Educational Attainment	Relative Life Expectancy	Relative Expenditure on Education	Relative Expenditure on Health	Relative Taxes	Log(Migrant stock)	Relative FDI	Trade	Multilateral Resiatance
Relative Youth Unemployment	1									
Relative Educational Attainment	-0.3624	1								
Relative Life Expectancy	-0.3731	0.1031	1							
Relative Expenditure on Education	-0.1471	0.5048	0.1815	1						
Relative Expenditure on Health	-0.4147	0.1799	0.8374	0.3974	1					
Relative Taxes	-0.1517	0.1165	0.3356	0.303	0.2337	1				
Log(Migrant stock)	0.1569	0.2287	-0.6065	0.1326	-0.4376	0.057	1			
Relative FDI	-0.0466	-0.1004	-0.1417	-0.059	-0.15	0.0118	0.0421	1		
Trade	-0.0431	0.1615	-0.4588	0.0173	-0.3706	0.1985	0.6732	0.0993	1	
Multilateral Resiatance	-0.041	0.1497	0.4268	0.3024	0.3754	0.1423	-0.0861	-0.0445	-0.0386	1

Appendix C

Dependent variable: Log(Emigration)

Table C.1 - Benchmark model (Hypotheses 1 and 2): Modified Wald test

Test Summary	x^2 Statistic	*p*-value
	5703.43	0.0000***

Notes: (i) Ho: sigma(i)^2 = sigma^2 for all i

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Baum (2001) discusses that the power of the Modified Wald test is low when N (number of countries) is large and T (years) is small. Therefore the likelihood ratio has also been tested.

Table C.2 - Benchmark model (Hypotheses 1 and 2): Likelihood-ratio test

Test Summary	x^2 Statistic	*p*-value
	454.19	0.0000***

Notes: (i) Ho: homosk nested in hetero (panels are homoscedastic).

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Table C.3 - Benchmark model (Hypotheses 1 and 2): Pesaran's test

Test Summary	Statistic	*p*-value
	13.271	0.0000***

Notes: (i) Ho: cross-sectional dependence

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

(iii) Average absolute value of the off-diagonal elements = 0.425

Table C.4 - Benchmark model (Hypotheses 1 and 2): Wooldridge test

Test Summary	F-Statistic	*p*-value
	9.122	0.0037***

Notes: (i) Ho: no first-order autocorrelation

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

(iii) Average absolute value of the off-diagonal elements = 0.425

Dependent variable: Log(Emigration of nationals)

Table C.5 - Benchmark model (Hypotheses 1 and 2): Modified Wald test

Test Summary	x^2 Statistic	*p*-value
	59947.65	0.0000***

Notes: (i) Ho: sigma(i)^2 = sigma^2 for all i

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Table C.6 - Benchmark model (Hypotheses 1 and 2): Likelihood-ratio test

Test Summary	x^2 Statistic	*p*-value
	349.94	0.0000***

Notes: (i) Ho: homosk nested in hetero (panels are homoscedastic).

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Table C.7 - Benchmark model (Hypotheses 1 and 2): Wooldridge test

Test Summary	F-Statistic	*p*-value
	1.067	0.3059***

Notes: (i) Ho: no first-order autocorrelation

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

(iii) Average absolute value of the off-diagonal elements = 0.425

Dependent variable: Log(Emigration of non-nationals)

Table C.8 - Benchmark model (Hypotheses 1 and 2): Modified Wald test

Test Summary	x^2 Statistic	*p*-value
	13396.32	0.0000***

Notes: (i) Ho: sigma(i)^2 = sigma^2 for all i

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Table C.9 - Benchmark model (Hypotheses 1 and 2): Likelihood-ratio test

Test Summary	x^2 Statistic	*p*-value
	620.35	0.0000***

Notes: (i) Ho: homosk nested in hetero (panels are homoscedastic).

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

Table C.9 - Benchmark model (Hypotheses 1 and 2): Wooldridge test

Test Summary	**F-Statistic**	***p*-value**
	51.231	0.0000***

Notes: (i) Ho: no first-order autocorrelation

*(ii) ***significant at 1%; **significant at 5%; *significant at 10%.*

(iii) Average absolute value of the off-diagonal elements = 0.425

Appendix D

Table D.1 – Regressions output for Log(Emigration) as dependent variable: Part 1

VARIABLES	(1) Benchmark 1(H. 1 & 2)	(2) Benchmark 2(H. 1 & 2)	(3) Augmented 1(H. 3)	(4) Augmented 2.1(H. 3)	(5) Augmented 2.2(H. 3)	(6) Augmented 3 (H. 3)	(7) Augmented 4(H. 3)
Log(Relative GDP pc)	-1.863***	-1.863***	-1.652***	-1.733***	-1.824***	-0.799***	-0.970***
	(-10.18)	(-10.18)	(-8.390)	(-8.275)	(-9.192)	(-3.393)	(-3.835)
Log(Distance)	28.71***	-6.112***	-5.049***	-5.991***	-7.294***	-4.893***	-5.540***
	(16.72)	(-12.97)	(-7.493)	(-8.887)	(-8.701)	(-5.665)	(-6.091)
Log(Population Spain*Population destination)	5.235***	5.235***	4.703***	5.227***	5.725***	4.825***	5.062***
	(18.91)	(18.91)	(12.66)	(14.08)	(13.21)	(10.70)	(11.01)
Relative Inflation	-0.00127	-0.00127	-0.000738	-0.000959	-0.004***	-0.000579	-0.000588
	(-1.466)	(-1.466)	(-0.827)	(-1.101)	(-3.784)	(-0.567)	(-0.577)
Multilateral Resistance			0.0133**	0.0133**	0.00117	0.0162***	0.0134**
			(2.189)	(2.167)	(0.223)	(2.592)	(2.092)
Relative Youth Unemployment					0.0761**		
					(2.164)		
Relative Unemployment				-0.0387**		0.00226	0.00219
				(-2.308)		(0.106)	(0.104)
Language	-43.98***						
	(-16.15)						
Relative Inequality (Gini)						-0.105	-2.556**
						(-0.462)	(-2.175)
Square of Relative Inequality (Gini)							1.116**
							(2.132)
Constant	-309.5***	-45.21***	-45.58***	-48.03***	-45.42***	-51.07***	-48.49***
	(-16.83)	(-16.34)	(-16.62)	(-17.69)	(-14.75)	(-11.81)	(-10.88)
Observations	580	580	580	566	346	398	398
Number of countries	60	60	60	58	49	53	53
Country FE	YES	YES	YES	YES	YES	YES	YES

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) and Language are lagged; (iv) all models are estimated with country FE and FGLS.*

Table D.2 – Regressions output for Log(Emigration) as dependent variable: Part 2

VARIABLES	(8) Welfare 1 (H. 4&5)	(9) Welfare 2 (H. 4&5)	(10) Welfare 3 (H. 4&5)	(11) Welfare 4 (H. 4&5)	(12) Welfare 5 (H. 4&5)	(13) Relations 1 (H. 6)	(14) Relations 2 (H. 6)	(15) Relations 3 (H. 6)	(16) Relations 4 (H. 6)
Log(Relative GDP pc)	-1.419***	-0.887***	-1.220***	-1.183***	-1.570***	-1.559***	-1.216***	-1.247***	-1.260***
	(-4.140)	(-3.507)	(-5.601)	(-4.843)	(-4.651)	(-4.590)	(-4.678)	(-4.787)	(-5.027)
Log(Distance)	0.978***	-4.720***	-12.55***	-12.31***	9.253***	9.427***	-12.07***	-12.76***	-11.84***
	(4.127)	(-5.183)	(-14.32)	(-13.09)	(11.43)	(11.51)	(-12.87)	(-12.71)	(-12.76)
Log(Population Spain*Population destination)	6.490***	4.511***	6.275***	6.288***	6.388***	6.522***	6.444***	6.546***	6.168***
	(10.70)	(9.560)	(13.97)	(14.08)	(12.05)	(11.86)	(13.60)	(13.74)	(14.06)
Relative Inflation	-0.00214*	-0.000354	-0.00139	-0.00124	-0.003***	-0.0031***	-0.00179*	-0.00181*	-0.00193**
	(-1.671)	(-0.348)	(-1.509)	(-1.365)	(-2.695)	(-2.724)	(-1.743)	(-1.748)	(-2.056)
Multilateral Resistance	0.00110	0.0170***	0.0127*	0.0125*	0.00492	0.00303	0.00882	0.00596	0.00150
	(0.166)	(2.694)	(1.919)	(1.885)	(0.649)	(0.381)	(1.275)	(0.861)	(0.224)
Relative Unemployment	0.0250	0.0103	-0.0186	-0.0205	0.0135	0.0118	-0.0167	-0.0148	-0.00975
	(1.010)	(0.495)	(-0.907)	(-1.002)	(0.537)	(0.473)	(-0.796)	(-0.704)	(-0.485)
Relative Inequality (Gini)	-3.112***	-2.165*	-5.298***	-5.111***	-5.759***	-5.904***	-5.393***	-5.377***	-5.056***
	(-2.958)	(-1.832)	(-4.391)	(-4.167)	(-3.600)	(-3.746)	(-4.162)	(-4.125)	(-4.054)
Square of Relative Inequality (Gini)	1.069**	0.951*	2.220***	2.143***	2.388***	2.442***	2.254***	2.259***	2.216***
	(2.247)	(1.814)	(4.202)	(4.023)	(3.461)	(3.597)	(4.013)	(4.001)	(4.098)
Relative Educational Attainment	-0.167								
	(-1.097)								
Relative Life Expectancy		7.579***							
		(2.757)							
Relative Expenditure on Education			-0.192**	-0.214*	-0.126	-0.121	-0.231**	-0.219*	-0.297***
			(-2.048)	(-1.898)	(-0.717)	(-0.694)	(-2.015)	(-1.903)	(-2.660)
Relative Expenditure on Health				-0.0211	0.0334	0.0257	-0.0171	-0.0170	-0.0602
				(-0.515)	(0.660)	(0.526)	(-0.423)	(-0.411)	(-1.352)
Relative Taxes					0.0839	0.0908			
					(0.927)	(0.986)			
Log(Migrant Stock)						0.00522	0.0123	0.0148	0.0180
						(0.0551)	(0.138)	(0.165)	(0.206)
Relative FDI								0.00732	0.00876
								(1.180)	(1.530)
Trade									-0.0956***
									(-4.405)
Constant	-120.8***	-45.76***	-	-	-194.2***	-198.0***	-	-	-
	(-11.64)	(-10.21)			(-11.80)	(-11.56)			
Observations	161	398	299	299	271	271	299	299	299
Number of countries	31	53	45	45	40	40	45	45	45
Country FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) are lagged; (iv) all models are estimated with country FE and FGLS.*

Table D.3 – Regressions output for Log(Emigration of nationals) as dependent variable: Part 1

VARIABLES	(1) Benchmark 1(H. 1 & 2)	(2) Benchmark 2(H. 1 & 2)	(3) Augmented 1(H. 3)	(4) Augmented 2.1(H. 3)	(5) Augmented 2.2(H. 3)	(6) Augmented 3 (H. 3)	(7) Augmented 4(H. 3)
Log(Relative GDP pc)	-0.226	-0.140	-0.140	-0.174	-0.00171	-0.208	-0.267
	(-1.083)	(-0.589)	(-0.589)	(-0.714)	(-0.00525)	(-0.769)	(-1.054)
Log(Distance)	-1.061**	0.359	0.359	0.501	0.475	0.323	-0.0962
	(-2.261)	(0.975)	(0.975)	(1.306)	(1.084)	(0.627)	(-0.174)
Log(Population Spain*Population destination)	0.111	0.139	0.139	0.220	0.208	0.133	0.0963
	(0.906)	(0.983)	(0.983)	(1.542)	(1.417)	(0.822)	(0.614)
Relative Inflation	-0.0105**	-0.0105**	-0.0105**	-0.00887*	-0.0177**	-0.0130**	-0.0122**
	(-2.327)	(-2.312)	(-2.312)	(-1.837)	(-2.339)	(-2.271)	(-2.201)
Multilateral Resistance				0.00353	-0.0235	0.0106	0.0116
				(0.133)	(-0.647)	(0.273)	(0.304)
Relative Youth Unemployment				0.00321		0.0329	0.0806
				(0.0296)		(0.233)	(0.596)
Relative Unemployment	3.030***						
	(4.175)						
Language					0.334		
					(1.125)		
Relative Inequality (Gini)						-1.297	-14.94**
						(-1.151)	(-2.131)
Square of Relative Inequality (Gini)							6.301*
							(1.901)
Constant	9.366**	-1.653	-1.653	-4.898	-1.262	-1.387	9.479
	(2.479)	(-0.417)	(-0.417)	(-0.835)	(-0.181)	(-0.180)	(0.917)
Observations	563	563	563	549	344	386	386
Number of countries	60	60	60	58	53	55	55
Country FE	NO	NO	NO	NO	NO	NO	NO

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) and Language are lagged; (iv) all models are estimated with RE and FGLS.*

Table D.4 – Regressions output for Log(Emigration of nationals) as dependent variable: Part 2

VARIABLES	(8) Welfare 1 (H. 4&5)	(9) Welfare 2 (H. 4&5)	(10) Welfare 3 (H. 4&5)	(11) Welfare 4 (H. 4&5)	(12) Welfare 5 (H. 4&5)	(13) Relations 1 (H. 6)	(14) Relations 2 (H. 6)	(15) Relations 3 (H. 6)	(16) Relations 4 (H. 6)
Log(Relative GDP pc)	-0.354	-0.204	-0.165	-0.596	-0.648	-0.532	-0.528	-0.503	-0.526
	(-1.053)	(-0.560)	(-0.579)	(-1.287)	(-1.295)	(-0.934)	(-0.954)	(-0.905)	(-0.981)
Log(Distance)	0.255	-0.0977	-0.436	-0.426	-0.239	-0.265	-0.452	-0.455	-1.090
	(0.408)	(-0.176)	(-0.793)	(-0.748)	(-0.400)	(-0.441)	(-0.767)	(-0.773)	(-1.523)
Log(Population Spain*Population destination)	0.0199	0.101	0.0245	0.0552	0.0288	-0.0384	0.0161	0.0187	0.171
	(0.0834)	(0.631)	(0.146)	(0.311)	(0.148)	(-0.150)	(0.0665)	(0.0761)	(0.735)
Relative Inflation	-0.0167	-0.0123**	-0.0142**	-0.0144**	-0.0159**	-0.0162**	-0.0145**	-0.0144**	-0.0145**
	(-1.287)	(-2.187)	(-2.165)	(-2.221)	(-2.311)	(-2.335)	(-2.221)	(-2.167)	(-2.157)
Multilateral Resistance	-0.0148	0.0119	-0.0405	-0.0347	-0.0185	-0.0212	-0.0362	-0.0409	-0.0464
	(-0.278)	(0.313)	(-0.930)	(-0.795)	(-0.411)	(-0.466)	(-0.819)	(-0.949)	(-1.085)
Relative Unemployment	-0.00330	0.0824	0.140	0.125	-0.0220	-0.0262	0.122	0.137	0.131
	(-0.0217)	(0.599)	(0.851)	(0.772)	(-0.126)	(-0.150)	(0.755)	(0.851)	(0.799)
Relative Inequality (Gini)	-9.010	-14.75**	-21.29***	-20.16***	-19.89**	-18.93**	-19.66***	-19.18**	-18.15**
	(-0.918)	(-2.112)	(-3.081)	(-2.887)	(-2.303)	(-2.155)	(-2.686)	(-2.568)	(-2.575)
Square of Relative Inequality (Gini)	4.638	6.225*	9.430***	8.950***	8.964**	8.644**	8.788**	8.603**	8.070**
	(0.944)	(1.880)	(2.852)	(2.686)	(2.200)	(2.086)	(2.548)	(2.452)	(2.422)
Relative Educational Attainment	1.222**								
	(2.345)								
Relative Life Expectancy		-1.509							
		(-0.295)							
Relative Expenditure on Education			1.257**	1.264**	1.255**	1.196*	1.247*	1.267*	1.208*
			(1.983)	(1.986)	(1.961)	(1.819)	(1.892)	(1.886)	(1.801)
Relative Expenditure on Health				0.257	0.223	0.222	0.253	0.246	0.275
				(1.324)	(0.972)	(0.969)	(1.285)	(1.204)	(1.311)
Relative Taxes					-0.0506	-0.0465			
					(-1.042)	(-0.932)			
Log(Migrant Stock)						0.0613	0.0336	0.0304	0.157
						(0.379)	(0.209)	(0.188)	(0.802)
Relative FDI								0.0446	0.0419
								(0.923)	(0.888)
Trade									-0.256***
									(-2.582)
Constant	7.611	9.296	22.22**	19.84*	16.94	17.78	20.45*	20.67*	22.47*
	(0.722)	(0.911)	(2.043)	(1.821)	(1.393)	(1.409)	(1.791)	(1.826)	(1.946)
Observations	165	386	293	293	265	265	293	293	293
Number of countries	37	55	47	47	41	41	47	47	47
Country FE	NO	NO	NO	NO	NO	NO	NO	NO	NO

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) are lagged; (iv) all models are estimated with RE and FGLS.*

Table D.5 – Regressions output for Log(Emigration of non-nationals) as dependent variable: Part 1

VARIABLES	(1) Benchmark 1(H. 1 & 2)	(2) Benchmark 2(H. 1 & 2)	(3) Augmented 1(H. 3)	(4) Augmented 2.1(H. 3)	(5) Augmented 2.2(H. 3)	(6) Augmented 3 (H. 3)	(7) Augmented 4(H. 3)
Log(Relative GDP pc)	-1.011**	-1.011**	-0.814*	-1.058**	0.262	-0.667	-0.891*
	(-2.298)	(-2.298)	(-1.696)	(-1.994)	(0.522)	(-1.464)	(-1.900)
Log(Distance)	22.06***	-3.788***	-2.723*	-4.389**	1.590	-3.557**	-4.691**
	(5.330)	(-3.394)	(-1.687)	(-2.457)	(0.817)	(-2.010)	(-2.485)
Log(Population Spain*Population destination)	3.851***	3.851***	3.316***	4.211***	0.940	3.802***	4.319***
	(5.825)	(5.825)	(3.732)	(4.270)	(0.916)	(3.804)	(4.150)
Relative Inflation	0.00270	0.00270	0.00326	0.00292	0.000230	0.00181	0.00158
	(1.420)	(1.420)	(1.624)	(1.421)	(0.110)	(0.865)	(0.729)
Multilateral Resistance			0.0133	0.0101	0.0152	0.0157	0.0120
			(0.944)	(0.657)	(1.100)	(1.182)	(0.893)
Relative Youth Unemployment				-0.0669*		-0.0552	-0.0543
				(-1.704)		(-1.126)	(-1.136)
Relative Unemployment						-0.441	-2.676
						(-0.886)	(-1.099)
Language							1.025
							(0.939)
Relative Inequality (Gini)					0.264***		
					(3.038)		
Square of Relative Inequality (Gini)	-32.64***						
	(-4.964)						
Constant	-235.6***	-39.36***	-39.69***	-42.76***	-30.20***	-42.27***	-41.14***
	(-5.357)	(-6.735)	(-6.842)	(-6.824)	(-5.515)	(-5.979)	(-5.657)
Observations	578	578	578	565	346	397	397
Number of countries	60	60	60	58	49	53	53
Country FE	YES	YES	YES	YES	YES	YES	YES

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) and Language are lagged; (iv) all models are estimated with country FE and FGLS.*

Table D.5 – Regressions output for Log(Emigration of non-nationals) as dependent variable: Part 2

VARIABLES	(8) Welfare 1 (H. 4&5)	(9) Welfare 2 (H. 4&5)	(10) Welfare 3 (H. 4&5)	(11) Welfare 4 (H. 4&5)	(12) Welfare 5 (H. 4&5)	(13) Relations 1 (H. 6)	(14) Relations 2 (H. 6)	(15) Relations 3 (H. 6)	(16) Relations 4 (H. 6)
Log(Relative GDP pc)	-0.634	-0.776*	-0.976	-1.197	-1.809**	-1.719**	-1.184	-1.181	-0.770
	(-0.771)	(-1.807)	(-1.479)	(-1.559)	(-2.234)	(-2.069)	(-1.547)	(-1.543)	(-0.975)
Log(Distance)	1.309**	-3.033*	-9.662***	-8.928***	11.57***	11.91***	-9.406***	-9.405***	-8.938***
	(2.364)	(-1.662)	(-3.235)	(-2.994)	(4.687)	(4.868)	(-3.113)	(-3.105)	(-2.902)
Log(Population Spain*Population destination)	4.495**	3.236***	4.912***	4.576***	6.166***	6.475***	4.825***	4.833***	4.489***
	(2.169)	(3.123)	(3.379)	(3.150)	(3.946)	(4.094)	(3.300)	(3.300)	(2.977)
Relative Inflation	0.00282	0.00147	-0.000755	-0.00136	-0.00194	-0.00266	-0.00200	-0.00191	-0.00138
	(0.679)	(0.716)	(-0.259)	(-0.472)	(-0.629)	(-0.849)	(-0.670)	(-0.636)	(-0.420)
Multilateral Resistance	0.00621	0.0210*	0.00855	0.00704	0.000151	-0.00489	0.00125	0.00174	0.00718
	(0.290)	(1.660)	(0.482)	(0.398)	(0.00844)	(-0.260)	(0.0664)	(0.0916)	(0.357)
Relative Unemployment	-0.0371	-0.0656	-0.0138	0.00808	-0.0367	-0.0316	0.0153	0.0145	0.0251
	(-0.481)	(-1.459)	(-0.228)	(0.136)	(-0.545)	(-0.465)	(0.258)	(0.242)	(0.410)
Relative Inequality (Gini)	-11.73**	-2.605	-4.676	-4.554	-3.359	-3.983	-5.168	-5.517	-5.049
	(-2.305)	(-1.175)	(-1.194)	(-1.167)	(-0.679)	(-0.792)	(-1.313)	(-1.396)	(-1.246)
Square of Relative Inequality (Gini)	4.979**	1.011	1.665	1.626	0.996	1.238	1.863	2.004	1.816
	(2.206)	(1.001)	(0.983)	(0.971)	(0.475)	(0.581)	(1.101)	(1.185)	(1.065)
Relative Educational Attainment	-0.0244								
	(-0.0749)								
Relative Life Expectancy		10.88**							
		(2.192)							
Relative Expenditure on Education			0.156	0.257	0.0798	0.0770	0.233	0.201	0.181
			(0.322)	(0.521)	(0.150)	(0.146)	(0.472)	(0.404)	(0.356)
Relative Expenditure on Health				0.0792	0.0398	0.0169	0.0699	0.0644	0.111
				(0.486)	(0.210)	(0.0881)	(0.421)	(0.386)	(0.665)
Relative Taxes					-0.0951	-0.0898			
					(-0.805)	(-0.773)			
Log(Migrant Stock)						0.0801	0.0869	0.0865	0.0955
						(0.567)	(0.602)	(0.601)	(0.576)
Relative FDI								0.000242	0.00158
								(0.0230)	(0.142)
Trade									0.226
									(1.049)
Constant	-84.39**	-35.60***	-	-	-211.9***	-220.8***	-	-	-
	(-2.386)	(-5.041)			(-4.237)	(-4.389)			
Observations	161	397	299	299	271	271	299	299	299
Number of countries	31	53	45	45	40	40	45	45	45
Country FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%; (iii) all variables except Log(Distance) are lagged; (iv) all models are estimated with country FE and FGLS.*

Table D. 6 - Robustness check for relative deprivation (Log(Emigration) as dependent variable)

VARIABLES	(1) Benchmark (H. 1 & 2)		(2) Augmented (H. 3)		(3) Welfare (H. 4&5)		(4) Relations (H. 6)	
	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006	From 2002	From 2006
Log(Relative GDP pc)	-1.863***	-1.079***	-0.758***	-0.167	-0.671***	-0.292	-0.733***	-0.424*
	(-10.18)	(-6.031)	(-3.264)	(-0.772)	(-2.815)	(-1.196)	(-3.113)	(-1.798)
Log(Distance)	-6.112***	-4.407***	-4.999***	-3.393***	-12.27***	-11.17***	-11.77***	-9.908***
	(-12.97)	(-7.960)	(-5.910)	(-4.187)	(-12.89)	(-10.68)	(-12.75)	(-9.699)
Log(Population Spain*Population destination)	5.235***	4.475***	4.917***	4.099***	6.020***	5.416***	5.866***	4.907***
	(18.91)	(13.67)	(11.04)	(9.626)	(13.13)	(10.42)	(13.34)	(10.23)
Relative Inflation	-0.00127	-0.000547	0.00168	0.00160	0.00379**	0.00266	0.00298*	0.00176
	(-1.466)	(-0.664)	(1.056)	(1.123)	(2.131)	(1.593)	(1.800)	(1.268)
Relative Unemployment			-0.000955	0.0537***	-0.0176	0.0437**	-0.00345	0.0436**
			(-0.0464)	(2.895)	(-0.851)	(2.187)	(-0.175)	(2.392)
Relative Inequality (S80/S20)			-0.138*	-0.110	-0.283***	-0.272***	-0.248***	-0.192***
			(-1.804)	(-1.503)	(-3.394)	(-3.329)	(-3.103)	(-2.695)
Square of Relative Inequality (S80/S20)			0.0464	0.0452	0.0945**	0.101**	0.0777*	0.0640
			(1.236)	(1.219)	(2.163)	(2.241)	(1.820)	(1.528)
Multilateral Resistance			0.0185***	0.0142**	0.0204***	0.0167***	0.0106	0.0156**
			(2.849)	(2.534)	(2.890)	(2.640)	(1.567)	(2.451)
Relative Expenditure on Education					-0.167	-0.211	-0.243**	-0.262*
					(-1.360)	(-1.473)	(-2.068)	(-1.897)
Relative Expenditure on Health					-0.0393	0.0776**	-0.0723*	0.0483
					(-1.028)	(1.974)	(-1.681)	(1.200)
Log(Migrant Stock)							0.0398	-0.0857*
							(0.466)	(-1.889)
Relative FDI							0.0122**	0.00981*
							(2.228)	(1.867)
Trade							-0.0851***	-0.0953***
							(-4.037)	(-5.087)
Constant	-45.21***	-45.56***	-52.44***	-50.03***	-	-	-	-
	(-16.34)	(-16.07)	(-12.87)	(-12.22)				
Observations	580	455	398	359	299	265	299	265
Country FE	YES	YES	YES	YES	YES	YES	YES	YES

*Notes: (i) z-statistics in parentheses; (ii) *** significant at 1%, ** significant at 5%, * significant at 1%;(iii) all variables except Log(Distance) are lagged; (iv) all models are estimated with country FE and FGLS.*