Universidad del País Vasco Euskal Herriko Unibertsitatea

EKONOMIA ETA ENPRESA FAKULTATEA
FACULTAD DE ECONOMÍA Y EMPRESA

DOUBLE DEGREE: BUSINESS ADMINISTRATION AND LAW

ACEDEMIC YEAR 2023/24

ANALYSIS OF THE EFFECT OF IMMIGRATION ON AVERAGE SALARIES FOR NATIVES IN SPAIN: A PSEUDO PANEL DATA APPROACH

Author: Ane de Aramburu Ibarra
Director: Vicente Núñez Antón

INDEX

FIGURE INDEX

TABLE INDEX

ABSTRACT

In this work, we analyze the evolution of the average hourly salary of native workers in Spain for the period 2002-2018. In the descriptive part, we summarize the information related to the salaries in Spain, concentrating on the analysis of the wages and the wage gap between female and male workers. In this part, the location of immigrants in the different autonomous communities and their origin is also discussed. Moreover, as analyzed in later sections, these factors may have an effect on the Spanish wages. The data for our analysis corresponds to the "The Quadrennial Salary Structure Survey", which is available at the Spanish National Institute of Statistics (INE) website (https://www.ine.es). This is a survey carried out every four years in Spain and, for this specific study, the surveys for the years 2002, 2006, 2010, 2014 and 2018 were analyzed. In the empirical part, we have motivated and defined specific pseudo-panels and propose a multiple linear regression method approach corrected for the possible heteroscedasticity and autocorrelation that may be present in the data. Our conclusions support the fact that the average hourly salary for Spanish workers varies depending on the gender of the worker as well as on the level of studies and working experience that the specific worker may have.

RESUMEN

En este trabajo analizamos la evolución del salario medio por hora de los trabajadores nativos en España para el periodo 2002-2018. En la parte descriptiva, resumimos la información relacionada con los salarios en España, así como la evolución de los salarios y la brecha salarial entre hombres y mujeres. En esta parte, también se analiza la comunidad autónoma en la que se encuentran los inmigrantes y su origen. Ambas variables son factores que pueden influir en los salarios de los españoles. La base de datos de la que se han extraído los datos para este trabajo se corresponde con "La encuesta cuatrienal de estructura salarial", disponible en el Instituto Nacional de Estadística (INE) página web (https://www.ine.es). Se trata de una encuesta que se realiza cada cuatro años en España y, para este estudio, hemos analizado las encuestas de los años 2002, 2006, 2010, 2014 y 2018. En la parte empírica, hemos definido pseudo-paneles específicos y proponemos un modelo de regresión lineal múltiple corregida por la posible heteroscedasticidad y autocorrelación que pudiese existir en los datos. Nuestras conclusiones avalan el hecho de que el salario medio por hora de los trabajadores españoles varía en función del género del trabajador, así como de su nivel de estudios y su experiencia laboral.

LABURPENA

Lan honetan, Espainiako bertako langileen orduko batez besteko soldataren bilakaera aztertzen dugu 2002-2018 aldirako. Deskribapen-zatian, Espainiako soldatei buruzko informazioa laburbiltzen dugu, baita soldaten bilakaera eta gizonen eta emakumeen arteko soldata-arrakala ere. Zati honetan, etorkinak zein autonomia-erkidegotan dauden eta horien jatorria ere aztertzen da. Bi aldagaiek eragina izan dezakete espainiarren soldatetan. Lan honetarako datuak ateratzeko erabili den datu-basea "Soldata-egituraren lau urteko inkesta" da, Estatistikako Institutu Nazionalean (INE) eskuragarri dagoena (https://www.ine.es). Espainian lau urtean behin egiten den inkesta da, eta, azterlan honetarako, 2002, 2006, 2010, 2014 eta 2018. urteetako inkestak aztertu ditugu. Alderdi enpirikoan, pseudo-panel espezifikoak definitu ditugu, eta erregresio lineal multipleko eredu bat proposatzen dugu, datuetan egon daitekeen heteroszedastizitate eta autokorrelazio posibleak zuzenduta. Gure ondorioek bermatzen dute Espainiako langileen orduko batez besteko soldata aldatu egiten dela langilearen generoaren, ikasketa-mailaren eta lan-esperientziaren arabera.

1. Introduction

The main objective of this work is to study the effect of immigration on the wages of natives in Spain. This study will be carried out using econometric analysis techniques, understanding Econometrics as defined by prominent authors in this field such as Maddala (2001): "*Econometrics is the application of statistical and mathematical methods to the analysis of economic data, with the purpose of giving empirical content to economic theories and verifying or refuting them*".

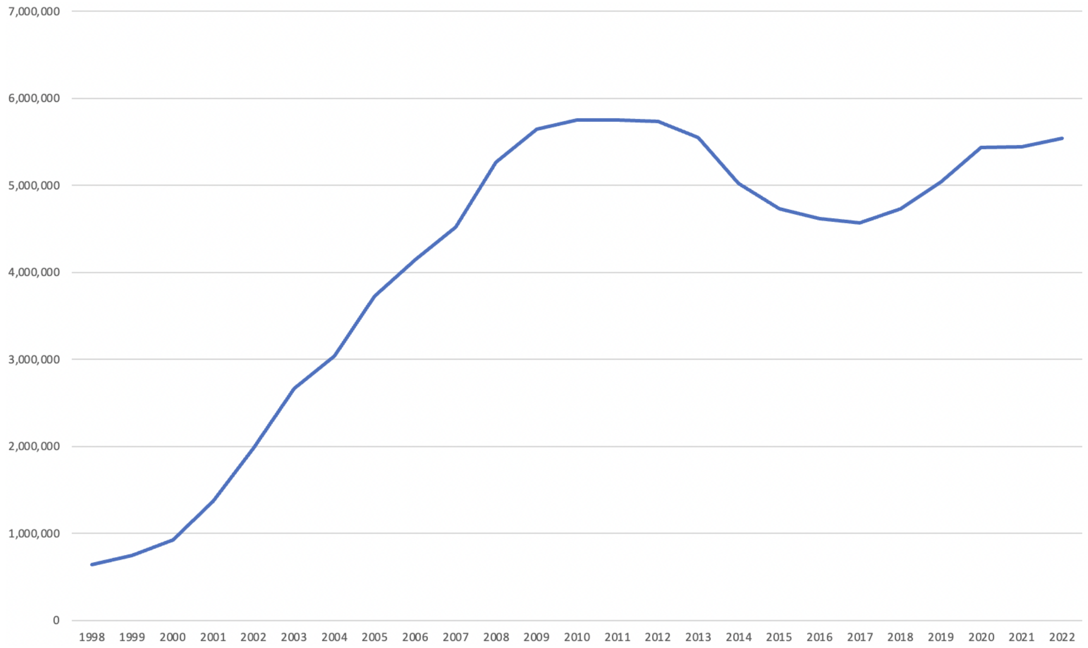
The effects of immigration on domestic labor markets has been a persistent and fiercely debated issue in many industrialized countries. In particular, there has been much debate among economists and policymakers alike as to whether immigration has a negative impact on native labour market outcomes (Gerfin, et al., 2010). We start by defining immigration so that it is clearly understood in the context of this work. We would like to mention that we concentrate on the study of the legal or regular immigration in Spain. That is to say, people of a nationality other than Spanish who have a valid residence and work permit (except in cases where this is not required such as, for example, foreigners from the European Union). In other words, this work does not cover all immigration in Spain, since immigrants in an irregular situation are not included in the database used to carry out the econometric analysis.

The reality is that there are very few empirical papers that study the phenomenon of immigration and the wages of the natives in Spain. A similar paper for this is Carrasco et al. (2008), where the authors studied the effect of immigration on the labor market of native workers in Spain on the 1990s. These authors used the population censuses of 1991 and 2001, work permits between 1993 and 1999 from the Ministry of Labor, and the Wage Structure Survey of 2002. They concluded that there was not a significant negative effect of immigration on the labor market and on the wages of natives in Spain. Moreover, their analysis was based on Borjas (2003), who investigated the same situation in the same decade, but in the United States. The economic analysis of immigration became increasingly relevant in the US political debate in the early 1980s, due not only to the high volume of inflows of foreign workers during the previous decade, but also to the transformations that the US labor market was beginning to undergo at that time: wage deterioration, growing inequality, union decline and increasing child poverty (Gonzalez, 2002).

Therefore, the empirical literature in Spain on this topic is very scarce and these studies were written several years ago. Consequently, one of the aims here is to try to shed some light on the changes and new insights related to this topic. In order to carry out this research, we have selected the periods between 2002 and 2018. An important difference between the analysis in Carrasco et al. (2008) and the present work is that they only used the 2002 Wage Structure Survey from the National Institute of Statistics (INE) and here we use the ones for the years 2002, 2006, 2010, 2014 and 2018. The main differences generated by using different datasets will be addressed later on.

From the available information from INE (2022), in Spain there are many immigrants that come from the European Union, where there is free movement between countries and the inhabitants of the members can work in Spain without any additional job permit (Atienza Montero et al., 2020). Therefore, we only consider workers with a nationality other than Spanish, whether they need a work permit or not, always bearing in mind that they are all in a regular situation. Moreover, as defined by the National Institute of Statistics, an immigrant is a person who was born in a foreign country, older than 15 years old and who has lived in Spain for at least one year. Figure 1 displays the increase of immigration in Spain since 1998 up to 2008, and then from 2017 and on.

Figure 1: Total number of immigrants in Spain from 1998 to 2022.

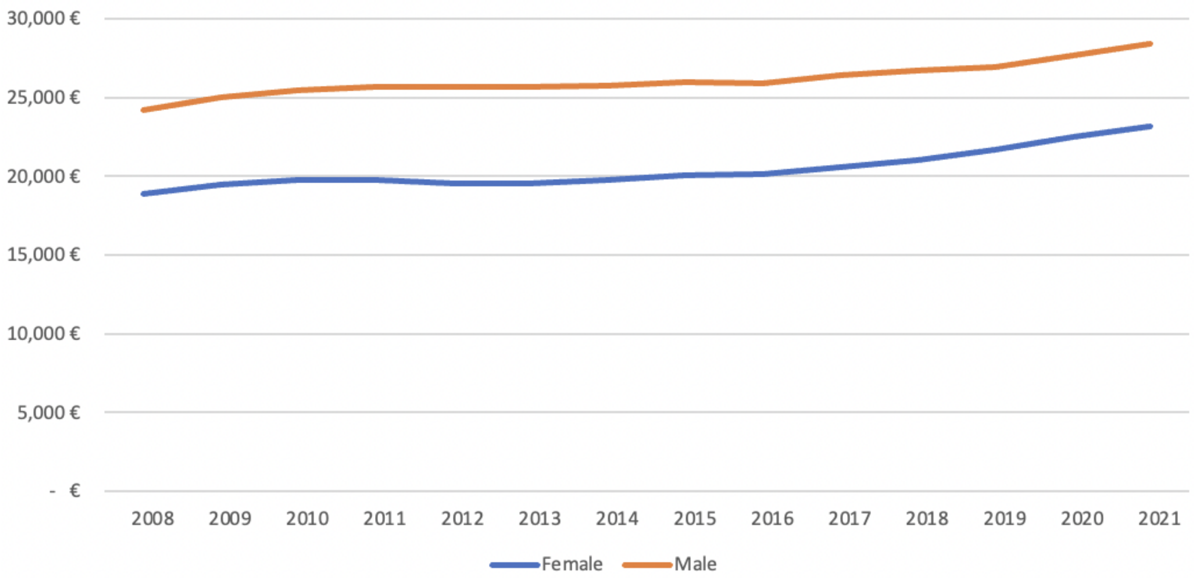


Source: *Own elaboration from INE (2022).*

Figure 1 shows how, after the economic crisis of 2008, immigration decreased and, afterwards, it increased again after 2017, when the economy started to recover, showing a steady increase later on. That is, Spain is one of the countries having a large number of immigrants in Europe (EUROSTAT, 2022).

Moreover, apart from the number of immigrants in Spain, we will also analyze the salary of workers in Spain, where changes in wages will be also studied in terms of gender differences. In order to have a general idea about the evolution of wages in this country over the years, Figure 2 includes the information on the changes of annual gross wage in Spain over the years.

Figure 2: Annual gross wage in Spain, males/females.



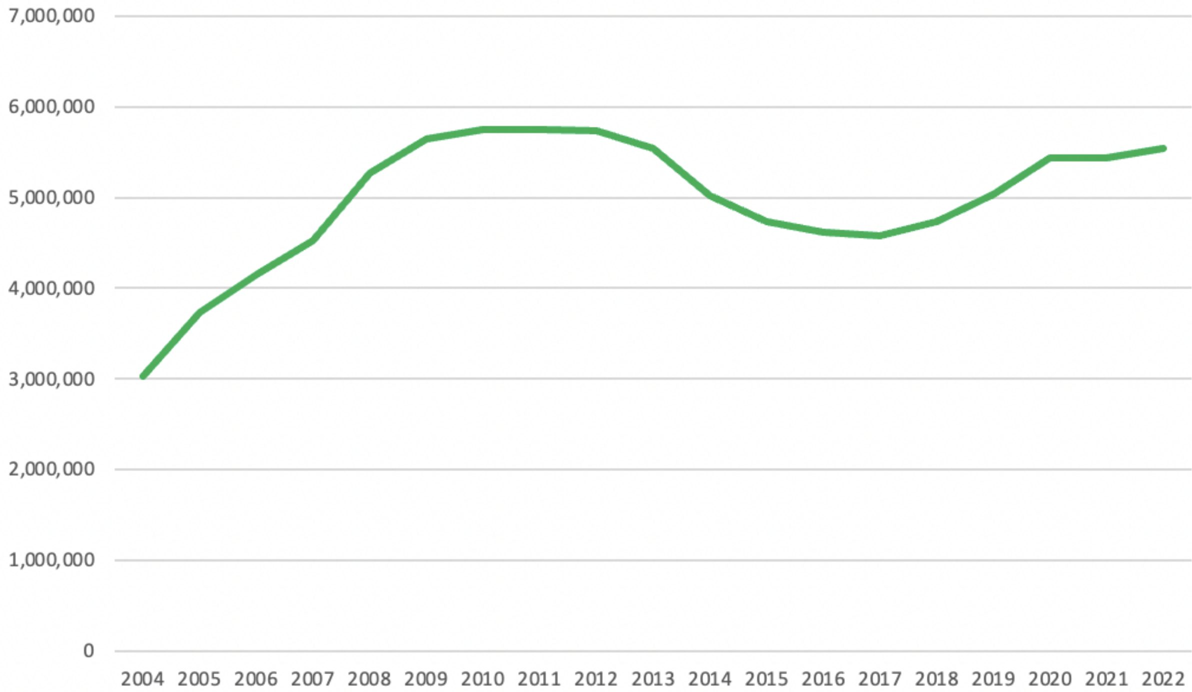
Source: *Own elaboration from data of INE (2021).*

From the information provided in Figure 2, we can conclude that women have always received much lower wages than men and that, in addition, wages have increased over the years, despite the small decreases in 2012, 2013 and 2014. Moreover, the general trend shown in wages for both males and females is quite similar.

2. Analysis of the Immigration in Spain.

We analyze the effect of immigration on the wages of natives in Spain for the period between 2002 and 2018. Therefore, having reliable information on the number of foreigners in Spain is going to be essential for this work. However, it should be made clear that, when analyzing the effect of foreigners on the wages of natives, only employed foreigners. That is, only those having a job are taken into account. Figure 3 shows the evolution of the foreign population in Spain from 2004 up to 2022.

Figure 3: Number of immigrants in Spain from 2004 up to 2022.

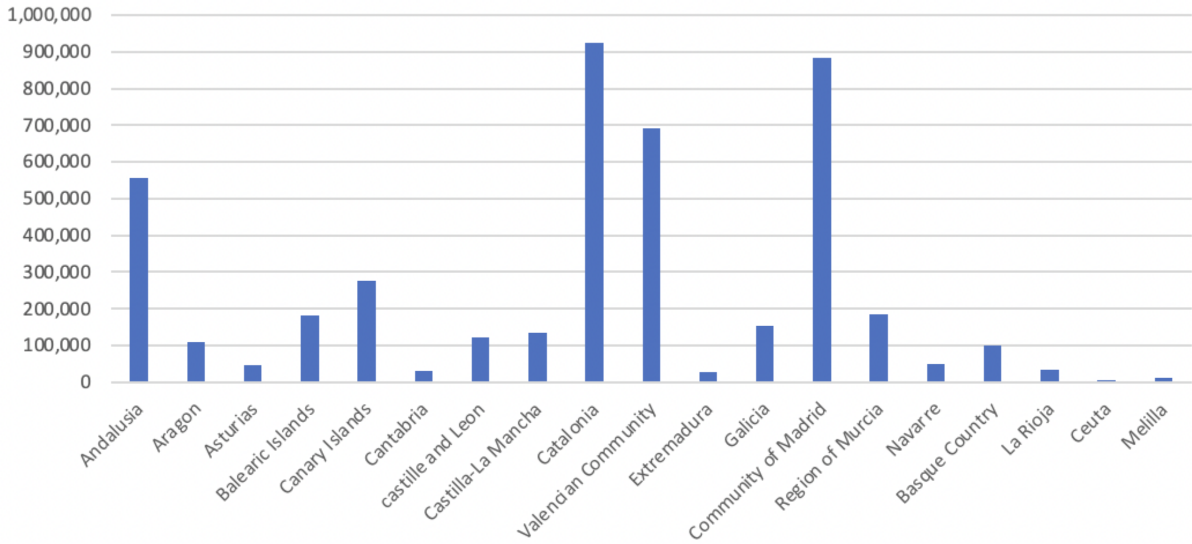


Source: *Own elaboration from the data of INE (2022).*

It can be seen how immigration did not grow after the economic crisis of 2008 and afterwards it began to decrease. Nevertheless, the immigration rose again after its recovery in 2017. It is well known how the crisis affected Southern Europe and Spain. From the available information, in 2022, the level of immigration has not exceeded the levels of shown between 2009 and 2012.

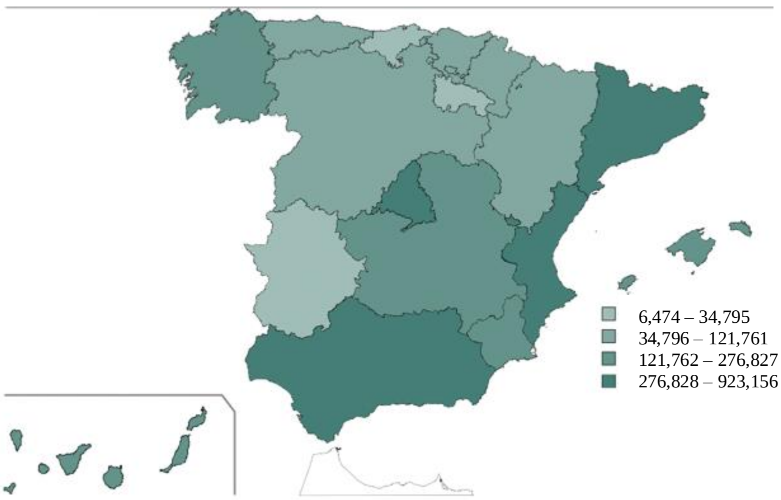
Moreover, it is important to highlight how immigration is usually concentrated in specific regions of a country. For example, in Spain, large cities such as Barcelona or Madrid have more employment offers and, thus, tend to have a larger presence of immigrants. Similarly, other Autonomous Communities that have important economic sectors may also have a notorious presence of immigrants. Figures 4 and 5 include information on the level of immigration for the different Autonomous Communities in Spain for the year 2022.

Figure 4: Number of immigrants by Autonomous Communities in Spain.



Source: *Own elaboration from the data of INE (2022).*

Figure 5: Distribution of immigrants in Spain.

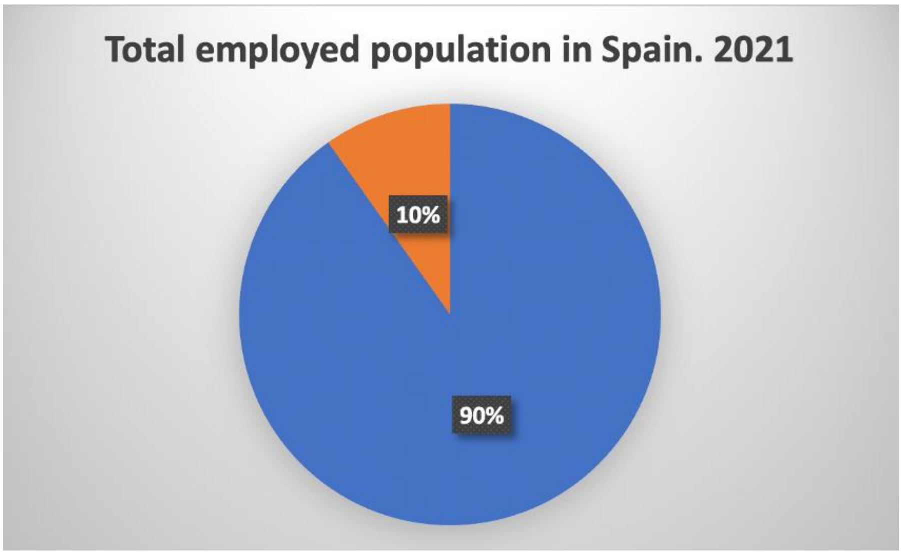


Source: O*wn elaboration from the data of INE (2022).*

Small regions such as Extremadura, Asturias, Cantabria, La Rioja and Navarre have fewer immigrants for various reasons: they do not have large cities and, in addition, they are far from the large cities such as Madrid, Barcelona, Valencia or Sevilla.

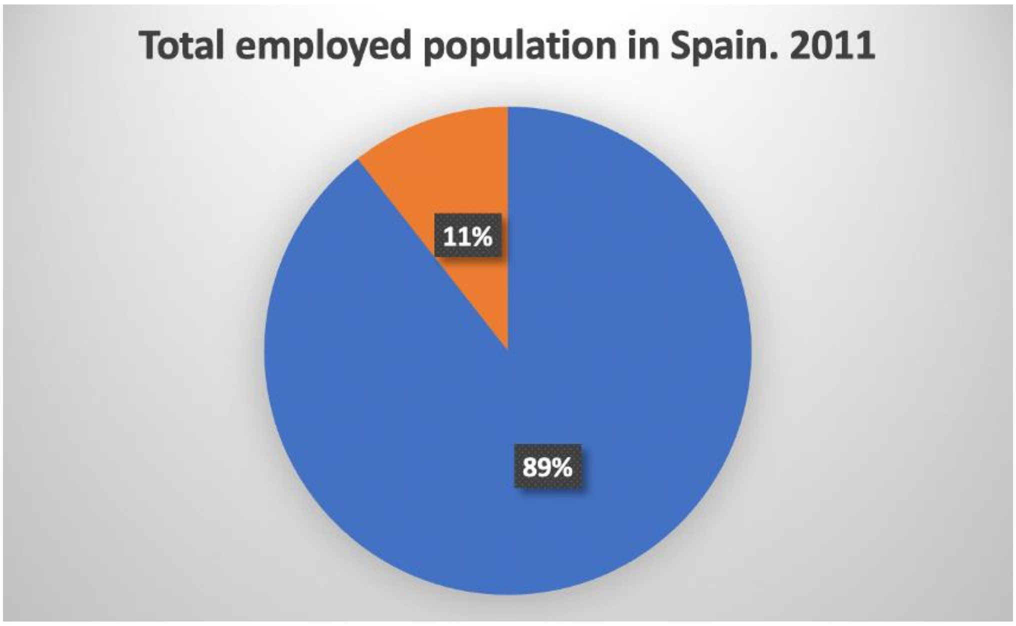
Immigration should be really analyzed using reliable data sources, since from the information included in the figures above it is not possible to compare percentages between foreign and native populations. The required data have been obtained from the INE Population Census, and it corresponds to information available every ten years. Figures 6 to 8 include information where we can compare the foreign and the Spanish population.

Figure 6: Total employed population percentages in Spain for immigrants and natives in 2021.



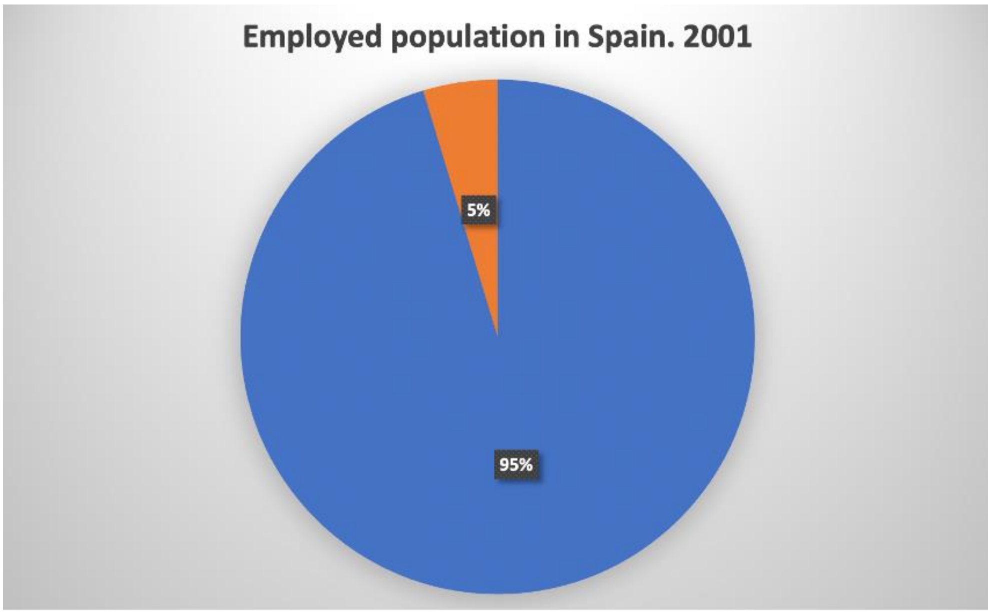
Source: *Own elaboration from the data of INE (2021).*

Figure 7: Total employed population percentages in Spain for immigrants and natives in 2011.



Source: *Own elaboration from the data of INE (2011).*

Figure 8: Total employed population percentages in Spain with for immigrants and natives in 2001.



Source: O*wn elaboration from the data of INE (2001).*

It is important to note that, in Figures 6 to 8, both populations (native and immigrant) are employed populations. From these figures, we can conclude that, since 2001, immigration has been increasing when compared to the native population, although from 2011 to 2021 it basically remains the same. In order to provide some additional information, data has been obtained from the Population Census and figures that will help us understand the real difference between employed native and immigrant populations are included in Table 1.

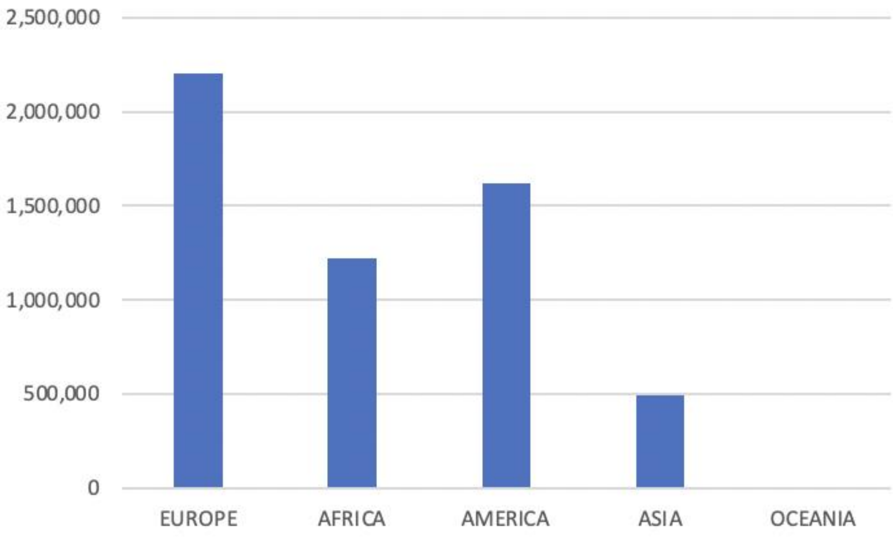
Table 1: Number of employed Spanish and immigrant populations in 2001, 2011 and 2021.

	TOTAL	SPANISH	IMMIGRANTS
2001	16,329,713	15,566,389	763,324
2011	17,514,550	15,658,920	1,855,635
2021	18,679,652	16,847,588	1,832,061

Source: *Own elaboration from the data of INE (2001, 2011 and 2021).*

In addition, one may come to the conclusion that most of the foreigners working in Spain come from third world countries, such as Africa, Eastern Asia or South America. However, Figure 9 includes some relevant information about the continent of origin of foreigners living in Spain in 2022.

Figure 9: Continent of origin for immigrants living in Spain in 2022.



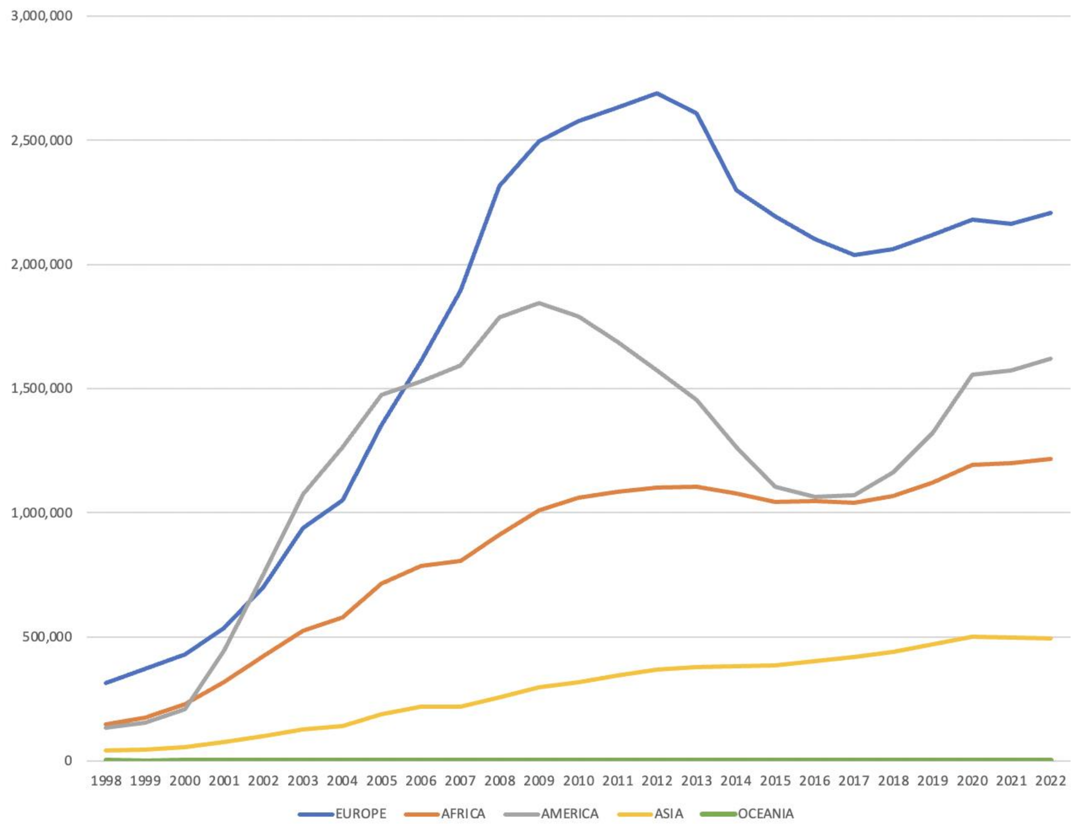
Source: O*wn elaboration from the data of INE (2022).*

As can be seen above, and coming as a surprise to many of us, Europe is the first continent of origin of Spain's immigrants (i.e., the entire European continent, not only the European Union), followed by America, Africa, Asia and Oceania.

We now analyze each continent and the countries with the highest presence in Spain. However, only Africa, Europe and South America are taken into account, because Central America, North America and Oceania have hardly any significant presence in Spain.

As for the evolution during the last 25 years, Europe has not always been the continent with the highest presence among immigrants in Spain. In the period between 2001 and 2006, America had the first place. Similarly, it can be observed how American immigrants decreased after the economic crisis of 2008 (see Figure 10).

Figure 10: Evolution of the number of immigrants, by continent, in Spain from 1998 up to 2022.

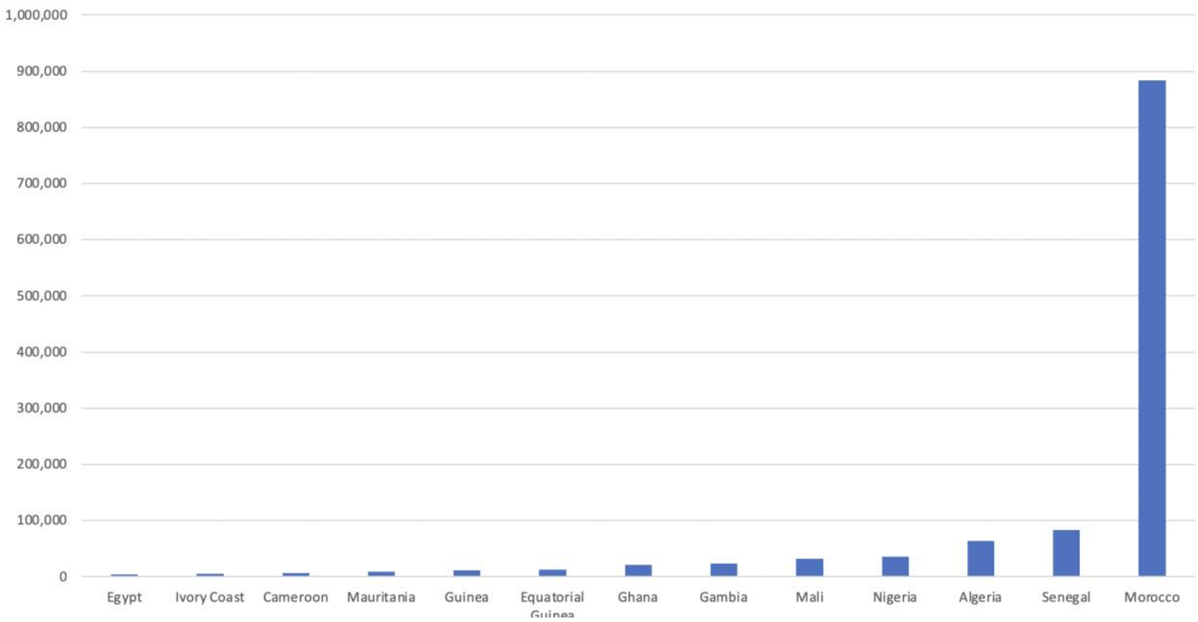


Source: *Own elaboration from the data of INE (2022).*

We now analyze the information for each continent in more detail, starting with Africa, and then following with Europe and South America.

Figure 11 includes the information for Africa, where it can be clearly seen that the vast majority of foreigners come from Morocco, with a large and very significant difference with regard to the other countries, respect to the following countries, which, in turn, are very close to each other.

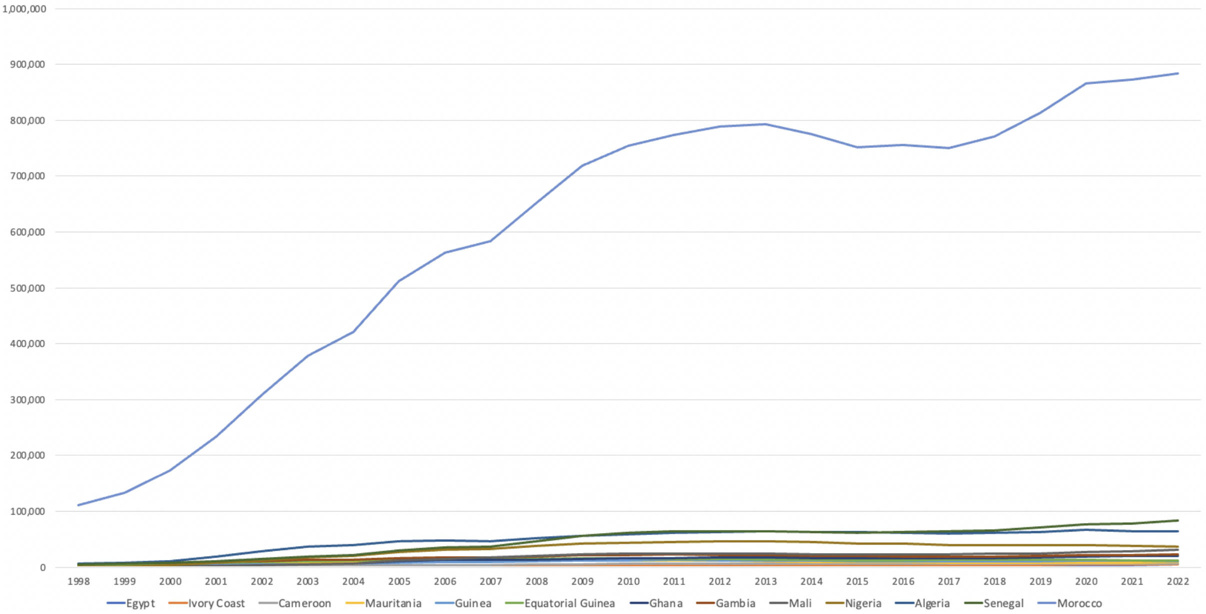
Figure 11: Country of origin of African immigrants in Spain for 2022.



Source: *Own elaboration from the data of INE (2022).*

Moreover, with regard to the evolution of immigration from this continent, we can see that it has been very similar over the years, except for the case of Morocco. As can be seen in Figure 12, for the year 1998, seen in 1998, figures for Morocco were not so far behind the rest of the remaining African countries. However, the difference in 2022 is a very significant one.

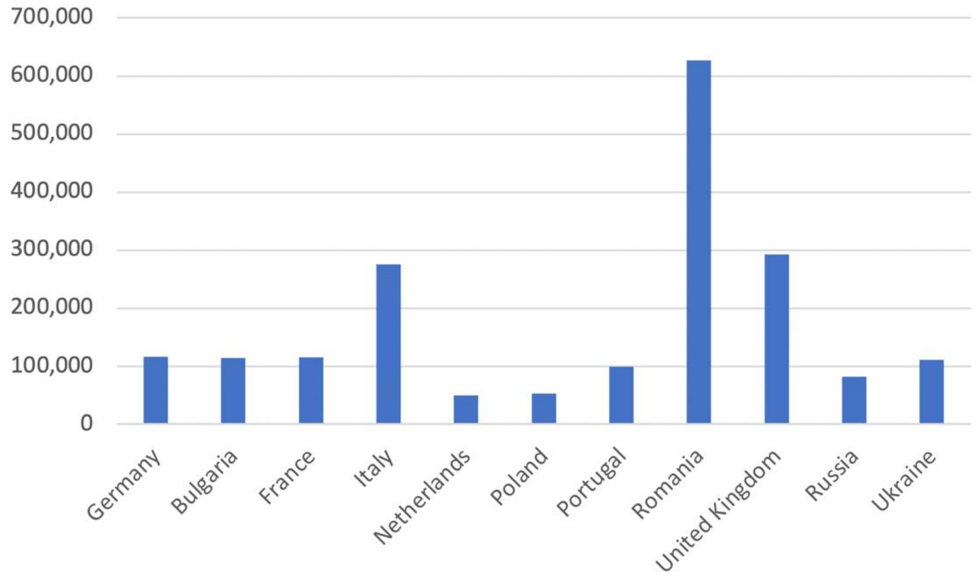
Figure 12: Evolution of the countries of origin of African immigrants from 1998 up to to 2022.



Source: *Own elaboration from the data of INE (2022).*

As we have already seen, Europe is the continent with the largest and most relevant presence for immigrants in Spain. Figure 13 includes the information for the different countries from the European Union with the largest number of immigrants in Spain. Among them, we have Romania, United Kingdom, Italy, Germany and France. However, it is clear that Romania is the country with the largest figures, followed by United Kingdom, but this difference is not as large as the one we had for African countries.

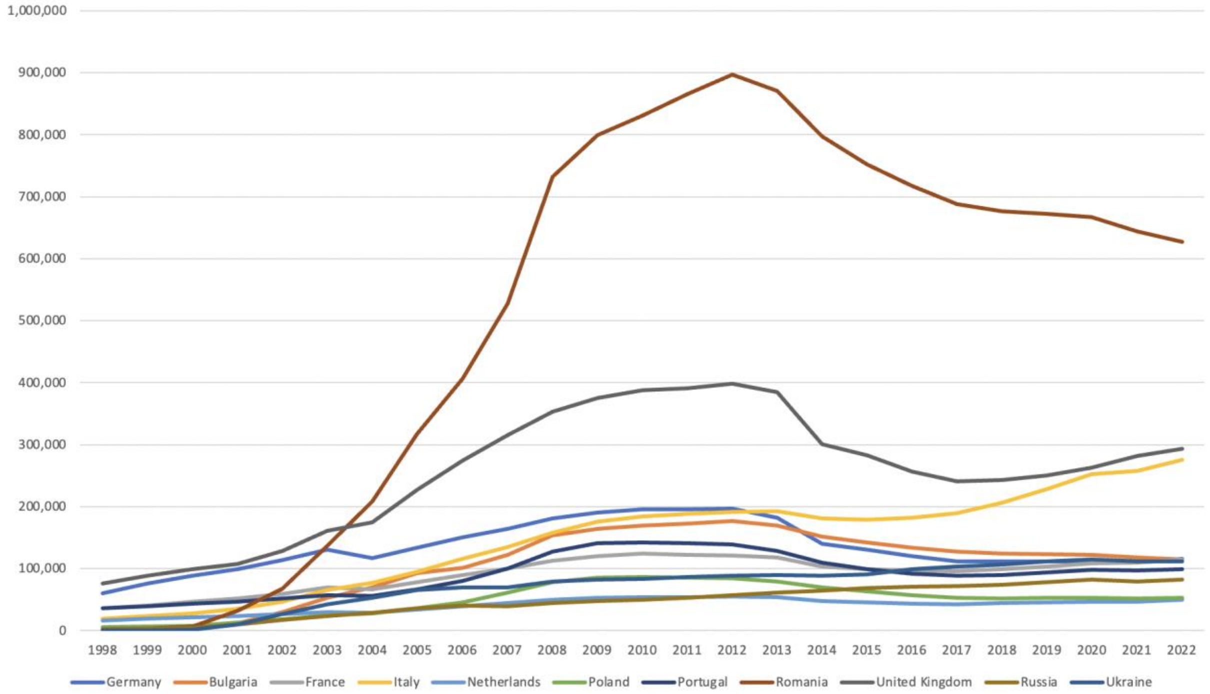
Figure 13: Country of origin for European immigrants in 2022.



Source: *Own elaboration from the data of INE (2022).*

However, the evolution over the previous 25 years has been changing. The case of Romania has not always been the same, it can be seen how until 2003 Romania was not the first country and that the United Kingdom and Germany surpassed this country. In turn, Romania had a peak in 2012 and since then it has not stopped decreasing, the opposite to the case of Italy, which has been increasing every previous year (see Figure 14).

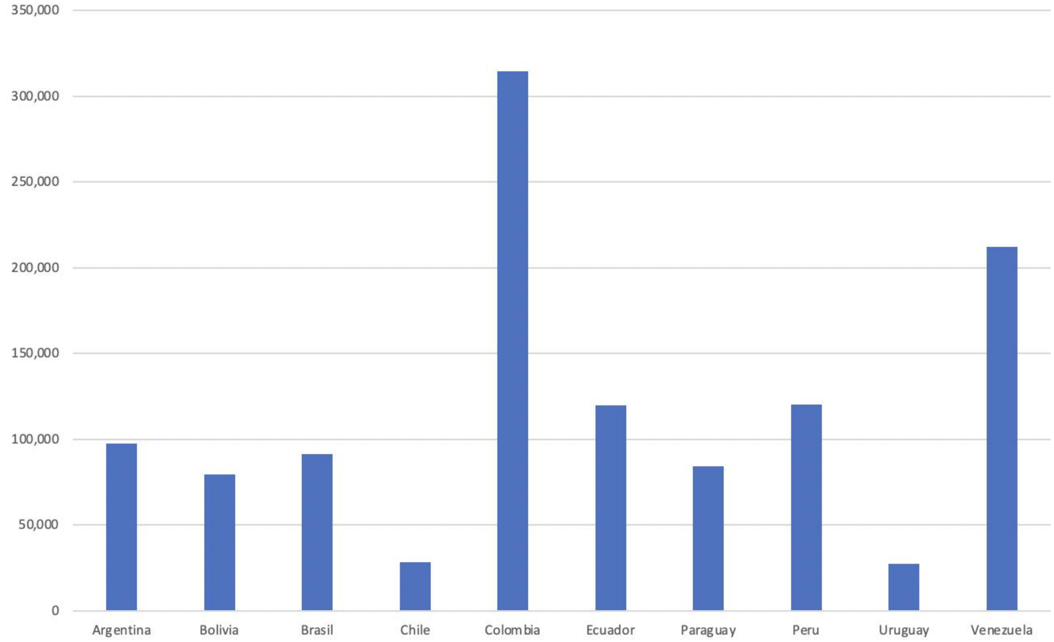
Figure 14: Evolution of the countries of origin for European immigrants in Spain from 1998 up to 2022.



Source: *Own elaboration from the data of INE (2022).*

Finally, with regard to the situation in South America, Figure 15 displays the information related to how the different countries may have a similar behavior in Spain for the year 2022. However, Colombia is the country with the largest number of immigrants in Spain, followed by Venezuela, Peru and Ecuador. As was already mentioned above, given their low presence in Spain in terms of immigrants, we are not going to analyze the cases of North and Central America.

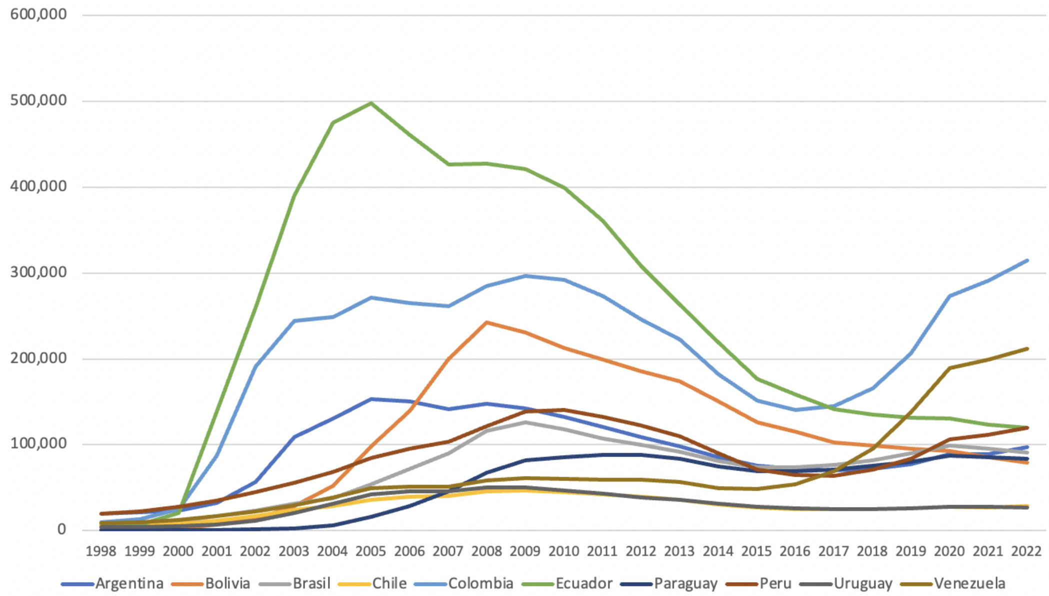
Figure 15: Country of origin of South American immigrants in 2022.



Source: *Own elaboration from the data of INE (2022).*

With regard to the evolution of immigrants from South America between 2000 and 2017, Ecuador was the first country of that continent with the most foreigners living in Spain. However, from 2017 onwards Colombia moved to that position. Moreover, it is also worth mentioning that Bolivia has had a steady decrease since 2008 (see Figure 16).

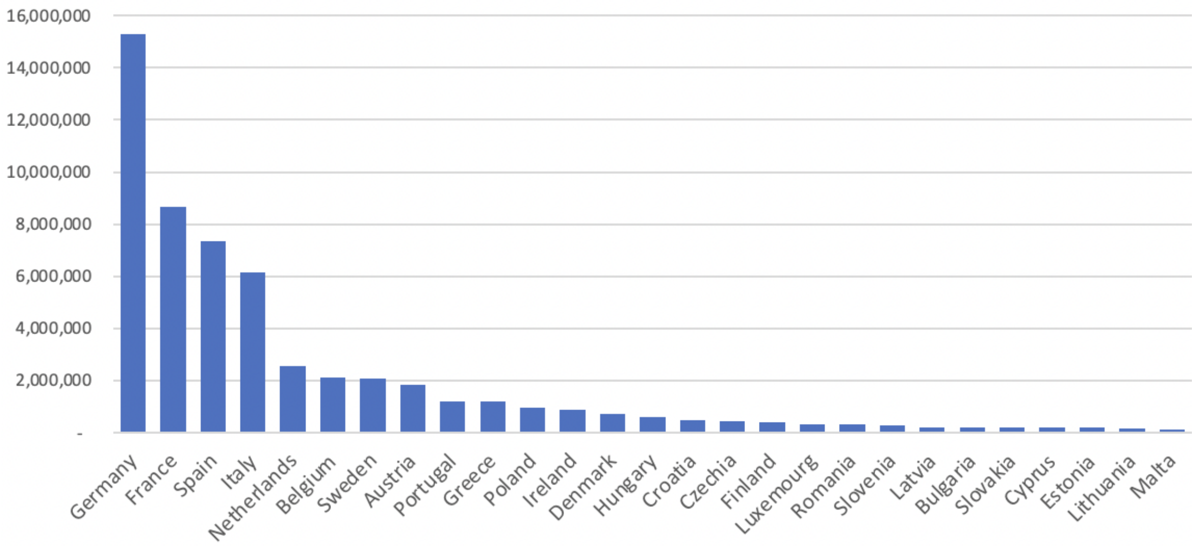
Figure 16: Evolution of countries of origin of South American immigrants from 1998 up to 2022.



Source: *Own elaboration from the data of INE (2022).*

Besides the immigration data for Spain, it would be also interesting to analyze the general situation in Europe, since the European Union has the same legislative norms on immigration and permits. Figure 17 includes information on the number of foreigners living in each European Union country, whether they are European or from outside Europe.

Figure 17: Number of immigrants in the European Union countries in 2022.

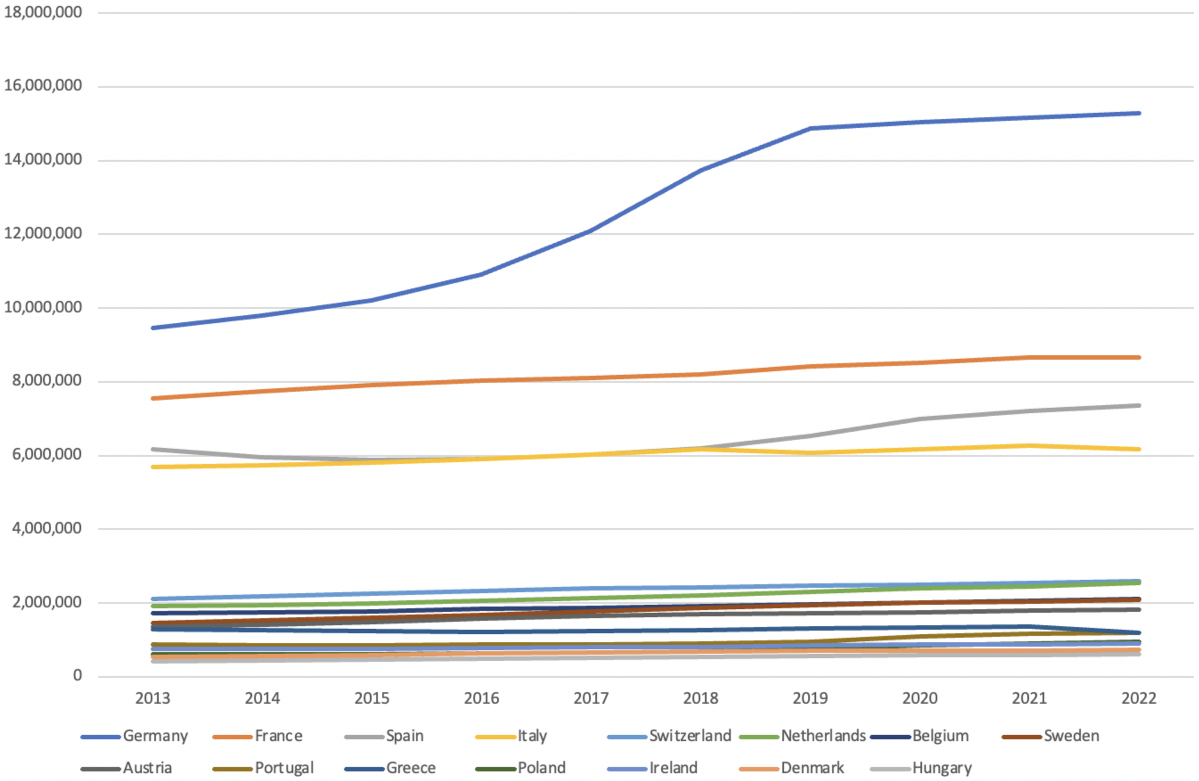


Source: *Own elaboration from data of EUROSTAT (2022).*

As can be seen in Figure 17, Spain is the third European country with the highest number of foreigners, right behind Germany and France. It should be noted that the United Kingdom is not part of the European Union since the 1st of February 2020. From the data available in EUROSTAT (2022), there is only information about the number of immigrants in the United Kingdom until 2019, which had a total of 9.469.015 immigrants in that year.

Figure 18 shows the evolution of the number of immigrants in the European Union countries. As can be seen, since 2018, Spain is on the third position after a few years having a similar number of immigrants as in Italy.

Figure 18: Evolution of the number of immigrants in the European Union countries in 2022.



Source: *Own elaboration from data of EUROSTAT (2022).*

An article from the International Monetary Fund (IFM) in 2016 (IMF, 2016) stated that: *"Migration, no matter how controversial politically, makes sense economically. A new IMF study shows that, over the longer term, both high- and low-skilled workers who migrate bring benefits to their new home countries by increasing income per person and living standards. High-skilled migrants bring diverse talent and expertise, while low-skilled migrants fill essential occupations for which natives are in short supply and allow natives to be employed at higher-skilled jobs. Moreover, the gains are broadly shared by the population. It may therefore be well-worth shouldering the short-term costs to help integrate these new workers".*

3. Analysis of Spanish Wages

We have already analyzed in detail the situation on the number of immigrants in Spain and Europe, and we now concentrate on the analysis of the wages, which will play the role of the dependent variable in our econometric analysis later on.

First of all, with regard to wages in Spain Table 2 includes information on the average wages classified by nationality and gender.

Table 2: Annual average wage in Spain classified by nationality and gender in 2020.

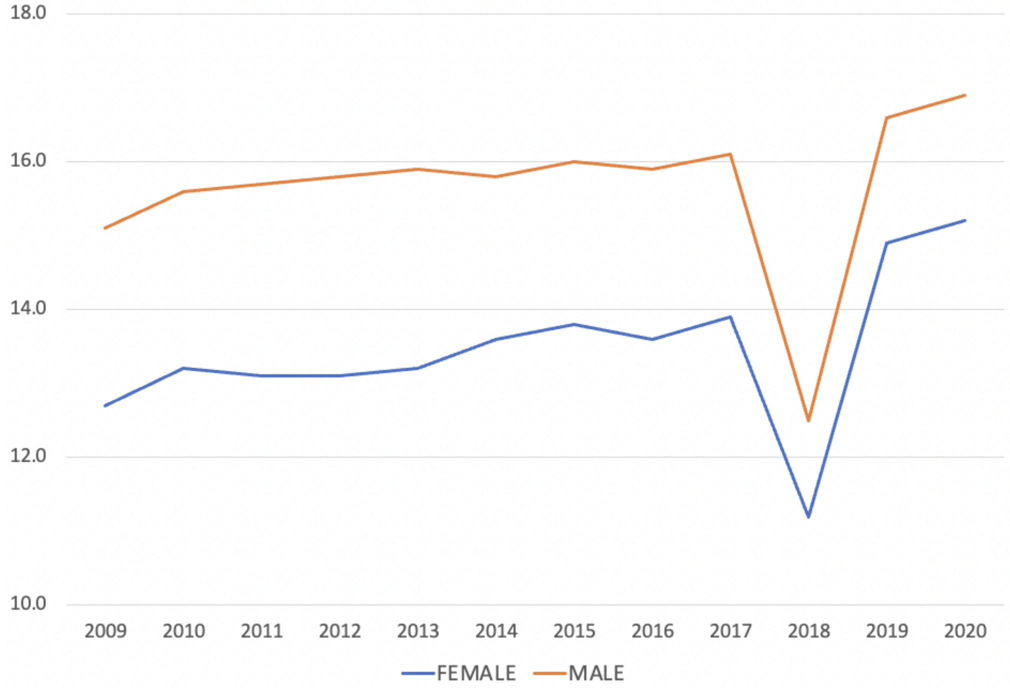
	Both genders	**Female**	**Male**
TOTAL	**25,165.51€**	**22,467.48€**	**27,642.52€**
Spanish	25,690.74€	22,894.30€	28,312.74€
EU without Spain	21,832.44€	18,741.52€	24,451.69€
Rest of Europe	19,852.61€	16,421.07€	23,677.28€
Africa	16,784.02€	13,840.89€	17,566.16€
America	16,100.85€	14,076.89€	17,929.05€
Rest of the world	16,433.04€	15,742.51€	16,774.91€

Source: *Own elaboration from data of INE (2020a).*

The highest salaries in Spain correspond to Spanish citizens, followed by foreigners from different EU countries and the rest of Europe. As was already seen, the vast majority of immigrants in Spain come from European countries. In addition, workers from Africa and America have the lowest salaries. Moreover, this table clearly shows the wage gap between males and females in Spain, which occurs in all of the nationalities included above.

Figure 19 shows the wage gap between male and female workers in Spain, without taking into account any additional characteristic for the specific workers. There has always been a wage gap. However, it is important to highlight how the salaries from both genders have the same tendencies, when one decreases or increases, the other gender decreases or increases as well.

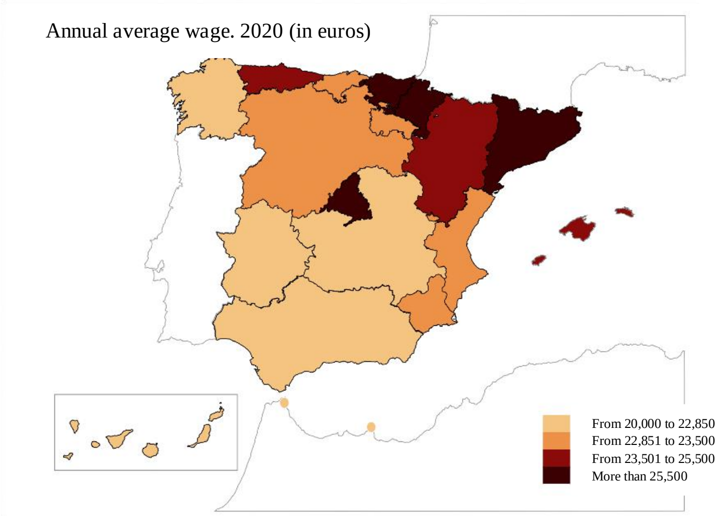
Figure 19: Evolution of the hourly average wage in Spain for male and female workers.



Source: *Own elaboration from data of INE (2020b).*

However, it is very important to clarify that salaries in Spain are not homogeneously distributed, as it is clear that salaries are higher in larger cities (see Figure 19). As we had already established in the precious section, immigration in Spain is mainly concentrated in Catalonia, Madrid, Andalusia and Valencia. Figure 19 includes the information that illustrates the issue of higher and lower salaries in the different Spanish Autonomous Communities.

Figure 20: Annual average wage in Spain in 2020 depending on the Autonomous Community.



Source: *Own elaboration from data of INE (2020a).*

Madrid, Catalonia, Navarre and the Basque Country have the highest salaries at the national level, followed by Aragon and Asturias. It does make sense to raise the issue that immigrants living in the largest cities, Madrid and Barcelona, have the highest wages. At the same time, living expenses in these cities are much higher than in the rest of the cities in Spain. Moreover, the Valencian and Andalusia Communities also have a large presence of immigrants but, in terms of wages, they are not as high as those in Madrid and Barcelona. This is mainly due to the fact that, the Andalusia and Valencian Communities are very large regions and, in addition, many provinces and towns live from agriculture and this activity does not usually involve having high salaries associated to it.

4. Analysis of Available Independent Variables

Table 3 shows some of the variables the INE provide, that include information on immigration and wages, and that have been selected to study the factors or variables the average wage of natives in Spain depends on.

Table 3: Characteristics of the working population.

Variable	Groups
- Gender	Male; Female
- Education	Primary; High School; Professional Training; Higher studies
- Years of experience	0-10; 11-20; 21-30; >30
- Year	2002; 2006; 2010; 2014; 2018
- Immigration rate	

Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

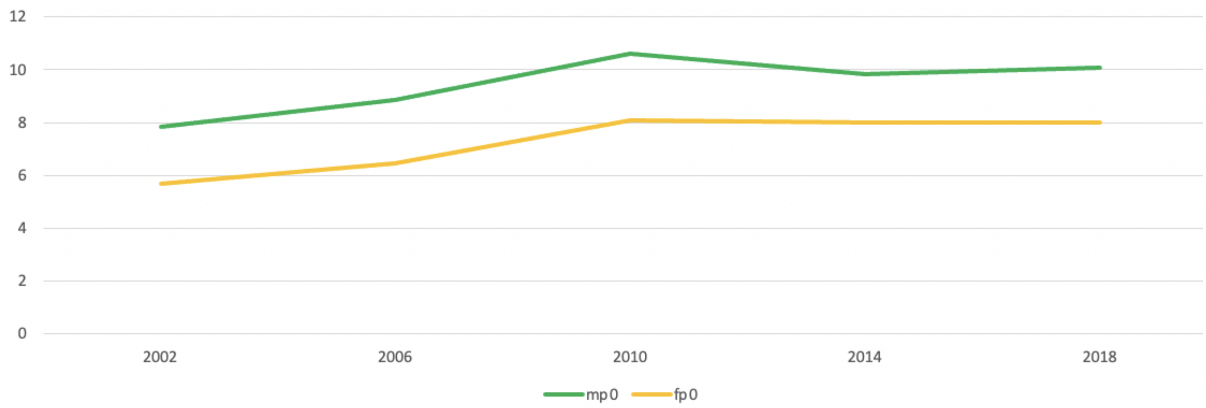
We do believe it is relevant to mention that, before directly analyzing the average wage of natives, it is important to know the context in which the different variables of interest influence this average hourly wage, and how it may change or on the different factors that may have an influence on it. In this data set, we have a set of different types of qualitative variables: gender, education, years of experience, year of the survey and year in which these workers are employed.

Analysis of the average hourly wage between men and women for different education level and years of experience for the years 2002-2018.

Figures 20 to 35 include information on the average salary of workers according to gender, education and years of experience for the period 2002-2018. These figure, for illustrative purposes, are broken down into sixteen different figures in order to be better able to assess, in a clearer way, the differences for men and women average salaries. It is paramount to say that in order to have the same data, the average wage is calculated with hourly average wage. This is due to the database available, where each individual had his own contract and different annual hours. As it was explained in the introduction, this work is using the Wage Structure Surveys from 2002 to 2018 and specifically the microdata databases have been used.

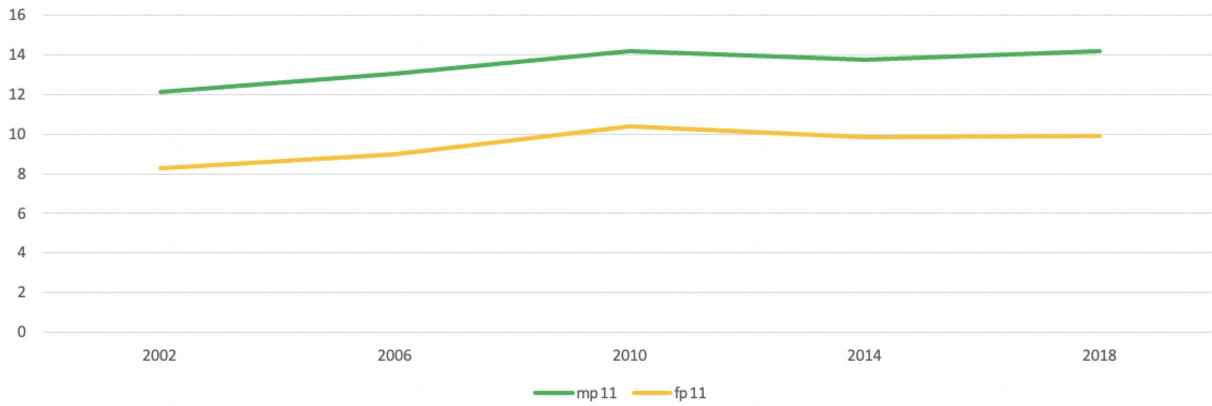
It is important to note that 32 groups were created (16 for each gender), so each group was given a name with *f* (feminine) or *m* (masculine) followed by the a letter (the education level) and a number (years of experience). For example, fhs11 means group of women with high school studies and 11-20 years of experience; mh0 means the group of males with higher education and 0-10 years of experience, and so on.

Figure 21: Evolution hourly average wages for the male and female groups with primary studies and 0-10 years of working experience.



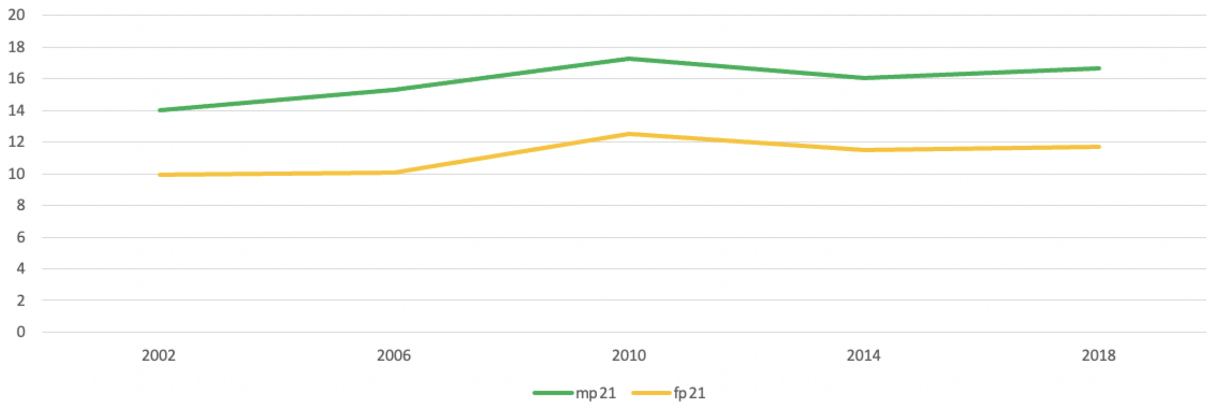
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 22: Evolution of hourly average wages for the male and female groups with primary studies and 11-21 years of working experience.



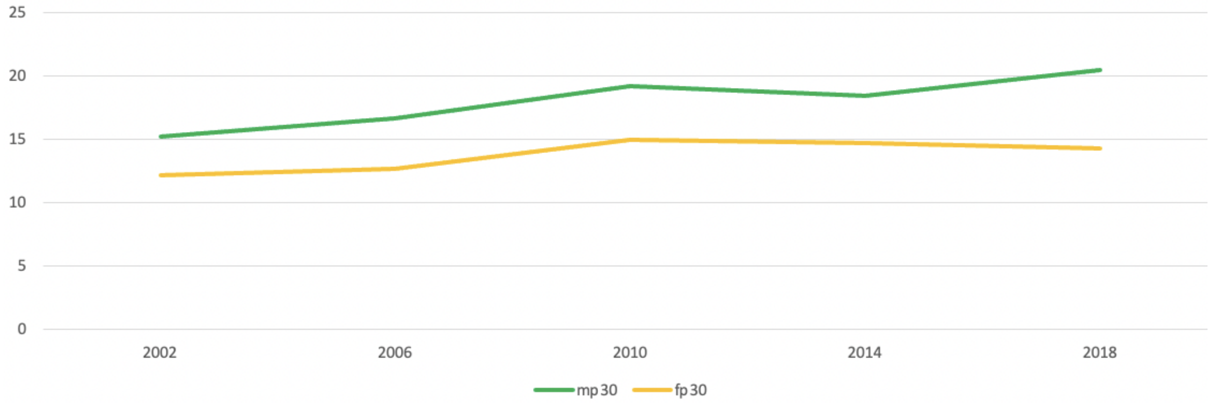
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 23: Evolution of hourly average wages for the male and female groups with primary studies and 21-30 years of working experience.



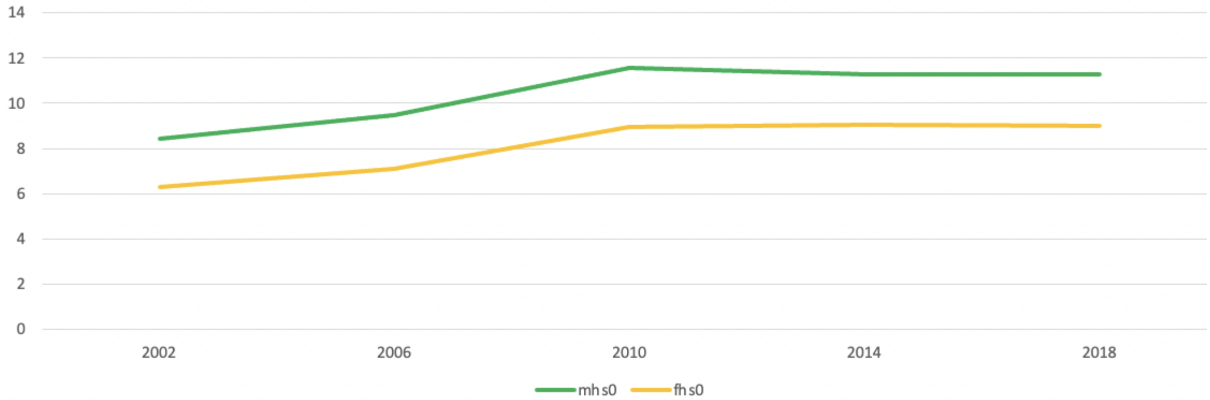
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 24: Evolution of hourly average wages for the male and female groups with primary studies and more than 30 years of working experience.



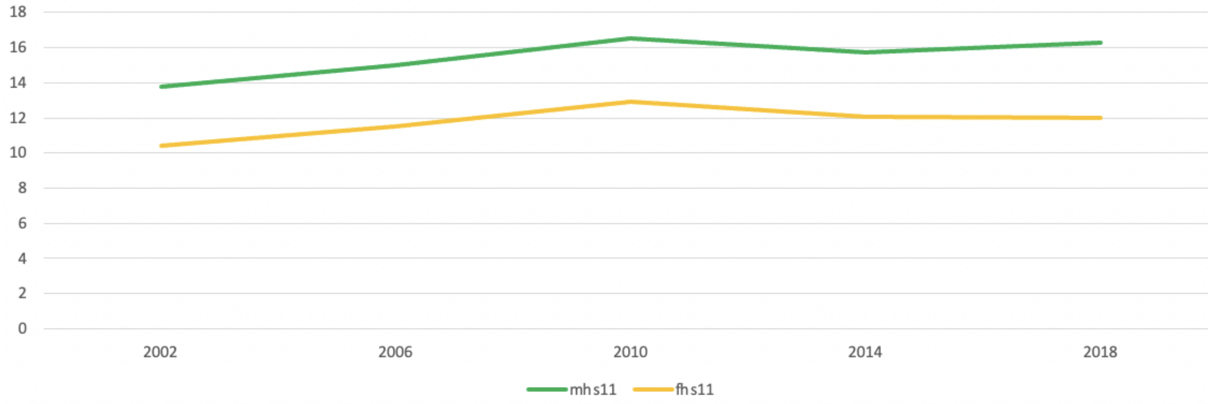
Source*: Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 25: Evolution of hourly average wages for the male and female groups with high school studies and 0-10 years of working experience.



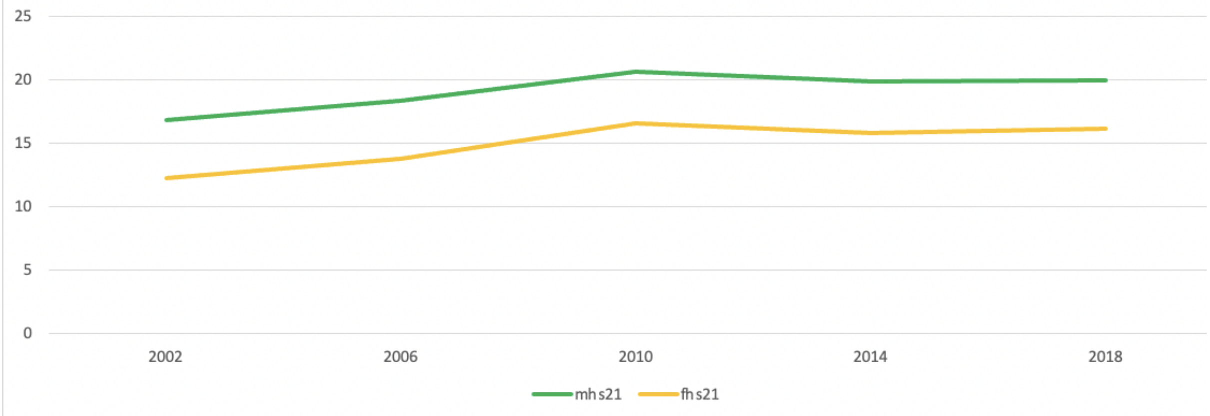
Source*: Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 26: Evolution of hourly average wages for the male and female groups with high school studies and 11-21 years of working experience.



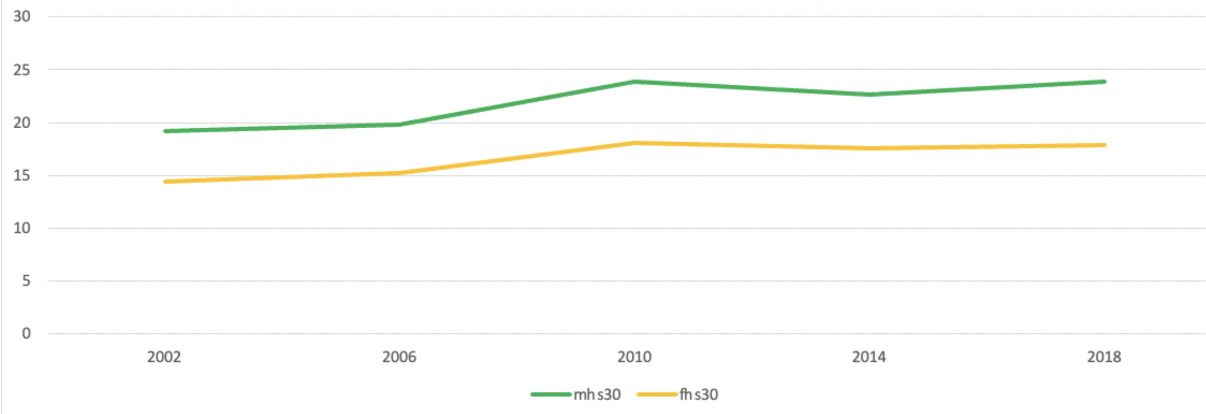
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 27: Evolution of hourly average wages for the male and female groups with high school studies and 21-30 years of working experience.



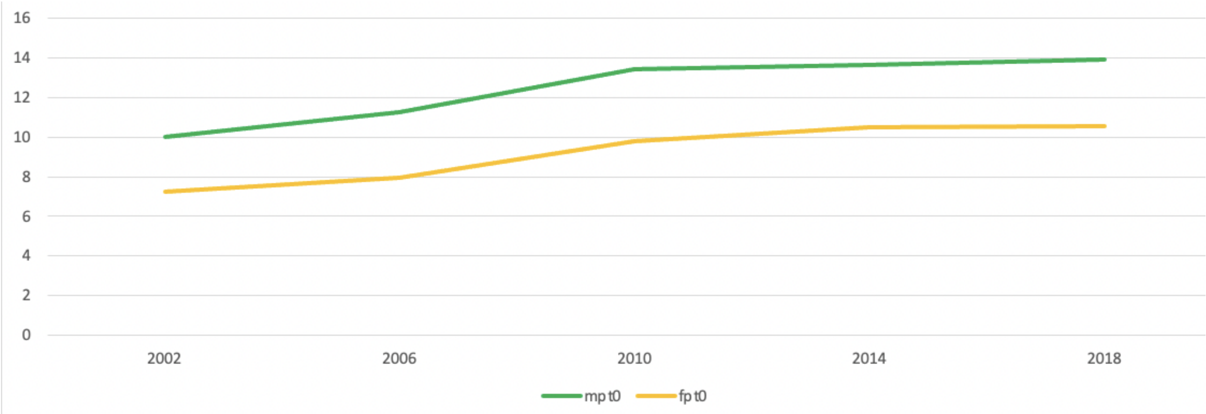
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 28: Evolution of hourly average wages for the male and female groups with high school studies and more than 30 years of working experience.



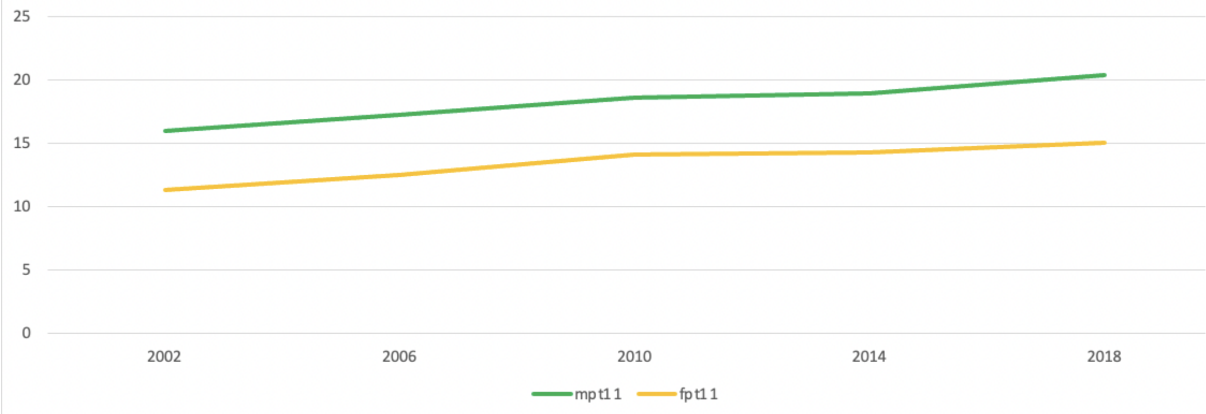
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 29: Evolution of hourly average wages for the male and female groups with professional training studies and 0-10 years of working experience.



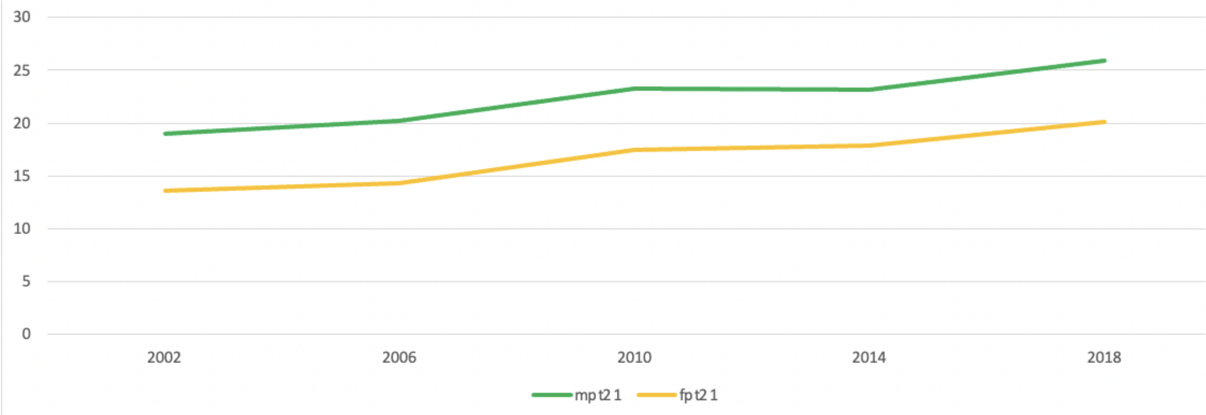
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 30: Evolution of hourly average wages for the male and female groups with professional training studies and 11-20 years of working experience.



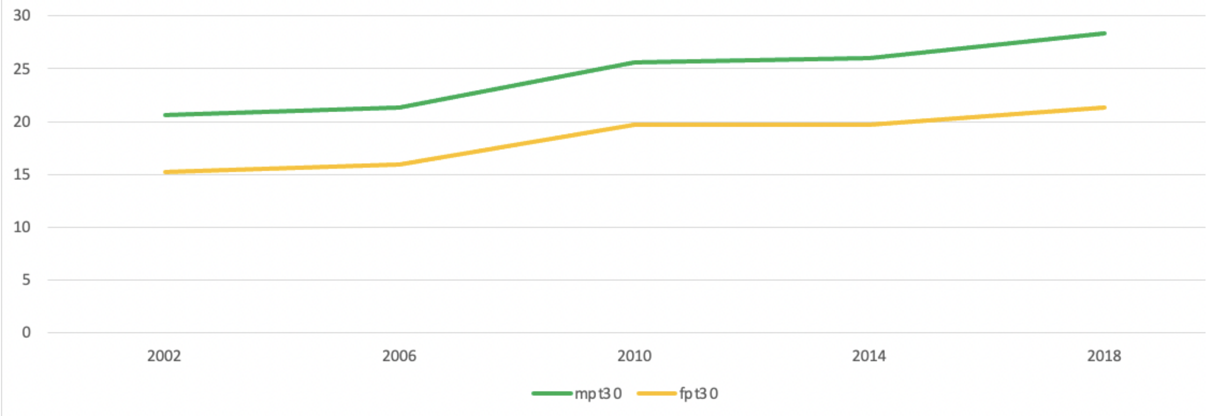
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 31: Evolution of hourly average wages for the male and female groups with professional training studies and 21-30 years of working experience.



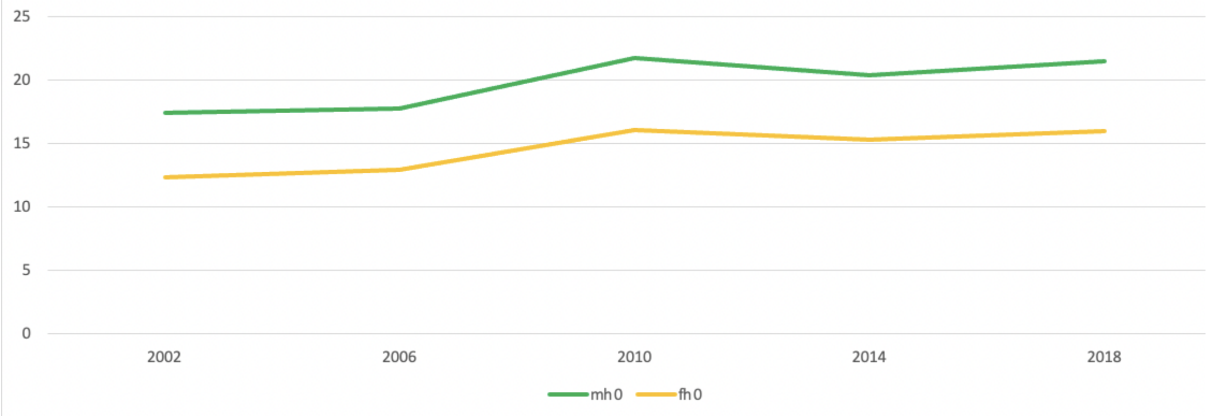
Source: *own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 32: Evolution of hourly average wages for the male and female groups with professional training studies and more than 30 years of working experience.



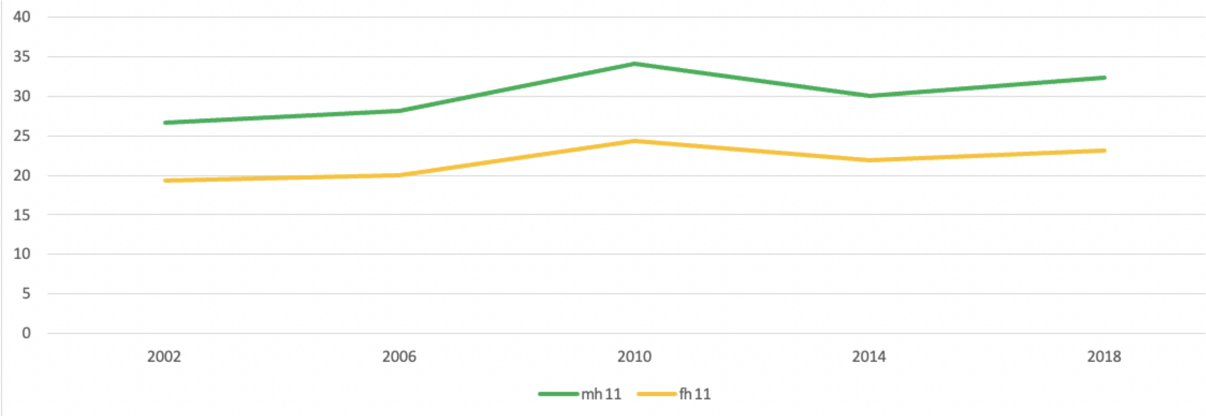
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 33: Evolution of hourly average wages for the male and female groups with higher studies and 0-10 years of working experience.



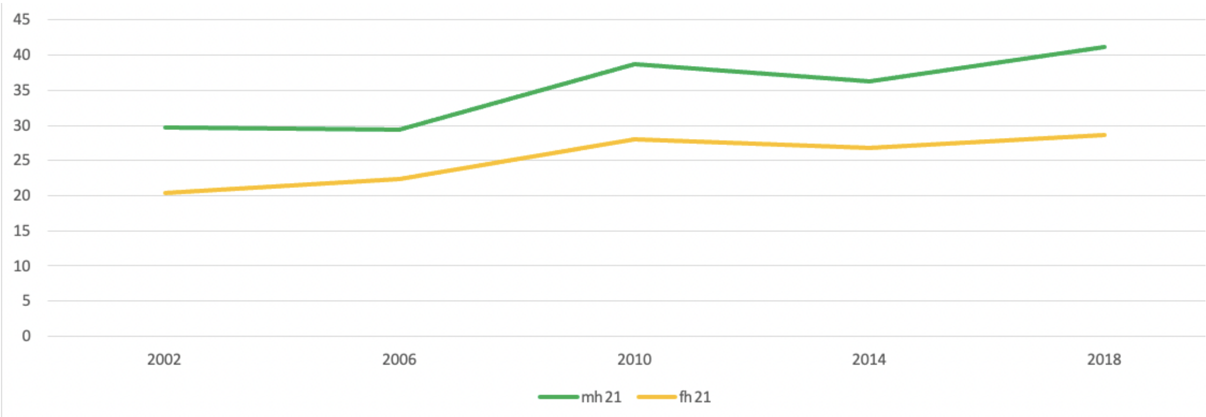
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 34: Evolution of hourly average wages for the male and female groups with higher studies and 11-20 years of working experience.



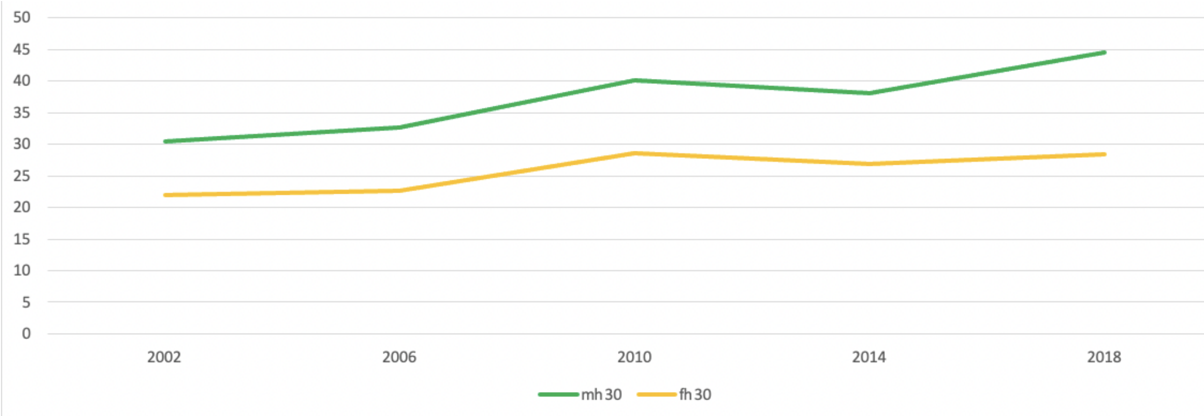
Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 35: Evolution of hourly average wages for the male and female groups with higher studies and 21-30 years of working experience.



Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

Figure 36: Evolution of hourly average wages for the male and female groups with higher studies and more than 30 years of working experience.



Source: *Own elaboration from data of INE (2002, 2006, 2010, 2014 and 2018).*

From the information included in Figures 20 to 35, we can observe that native male workers of any range of education and years of experience have a higher salary than native females. Similarly, wages for both gender groups have increased over the years, with the largest peak in the 2018 Wage Structure Survey. In addition, more experienced women (i.e., over 30 years) and with a higher education level (i.e., higher education) earn wages similar to those of men with the same education level but having minimum experience (see Figures 32 and 35). It is also clear that there is difference between the average hourly wages of workers with only primary education and workers with higher education. This latter assessment means that both education level and years of experience seem to be very relevant factors when it comes to hourly average wage. Moreover and for all the groups, it can be seen how, for the same group of education level and years of experience, male and female average hour wages follow the same tendency. That is, both increase or decrease at the same time.

From the information above, we can clearly conclude that there seems to be a wage gap between the hourly average wages of native men and those of native women for all of the available groups.

Additionally, for a few groups; that is, those having a primary education level and between 0 and 10 years of experience, a primary education level and between 21 and 30 years of experience, a high school education level and between 0 and 10 years of experience, a high school education level and between 11 and 20 years of experience, a high school education level and between 21 and 30 years of experience (see Figures 20, 22, 24, 22 and 26) the peak occurred in 2010 and not in 2018. For the remaining groups, the hourly average wage was higher in 2018 when compared to 2010.

It should be noted that these data on average hourly wages only take into account the native working population, which means that foreigners are not included here.

Therefore, we can conclude that there is an increasing trend in the average wage of Spaniards, with a specific drop between in 2014.

5. Multiple Linear Regression Model: Methodological Proposal

5.1 Building the Data Pseudo-Panels

The Wage Structure Survey's microdata from INE, which is the data under study here, includes information on a set of variables for a series of individuals who remain in the sample for only sixteen years. However, the data used are quadrennial data, which means the microdata is only available every four years. Moreover, we do not have individual information and, instead, data available is for specific categories classified according to the available variables for this data: gender, nationality, number of annual hours, wage base, education level and years of experience. The set of pseudo panel cohorts will constitute the pseudo panel data that we will study in this section, and the data where the appropriate methodology within this context will be applied (Deaton, 1985).

Consequently, five surveys are going to be used for the analysis; that is, the ones for the years 2002, 2006, 2010, 2014 and 2018. Therefore, there is a configuration similar to that of a panel or a longitudinal data setting, except for the fact that there is no individual information and, as already mentioned above, and we only have information on characteristics that group individuals together. Moreover, individuals on the different years are not the same and, in addition, some of them may have stayed in the survey or they may have been replaced by new individuals.

In sum, as individuals may be different for the different available years, in order to be able to estimate a given regression model within this context of linear models, there is the need to build or create an appropriate pseudo-panel data set, which defines the so-called pseudo-individuals based on common characteristics that they have and that can be observed over time. Additionally, different cohorts of pseudo-panels will be defined or built, so that the researcher is able to study their behavior for the different time-periods under study. At the end, the set of different pseudo panel cohorts will result in the pseudo panel data that we will study in this section.

In this specific case of immigration and wages of natives, standard cohorts are defined according to the variables time (year of the Wage Structure Survey), gender, education and years of working experience. In addition to these variables, there is another variable that is also partially available in the dataset under study and that we have modified accordingly to be able to incorporate it into the proposed model: the immigration rate. This rate was calculated dividing the number of immigrants for each group by the total population, i.e., natives and immigrants, of each group as well. In order to calculate this amount of people, the microdata was used due to the fact that allows us to make groups of people with the same characteristics. With regard to time, there are five available years: 2002, 2006, 2010, 2014 and 2018; as for gender, we have two available categories: males and females; as for education level, we have four available categories: primary education, secondary education, vocational training and higher education; and, finally, as for years of working experience, we have group them into four categories: 0-10, 11-20, 21-30 and more than 30 years. Therefore, we have a total of 160 observations or pseudo individuals for the different possible combinations, or pseudo panel cohorts, of time, gender, education level and working experience ($5 \times 2 \times 4 \times 4 = 160$ observations).

5.2 Multiple Linear Regression Model Adjusted for Heteroscedasticity and Autocorrelation.

The proposed model corrects for the possibility of heteroscedasticity and autocorrelation as originally described and introduced in Arellano (1987). Arellano’s corrections allow us to have a general structure with respect to heteroscedasticity and serial correlation. In our methodological proposal for the analysis of this pseudo-panel data, a multiple linear regression model is proposed, where the variables immigration rate, year, gender, education level and years of working experience are included as independent variables or covariates.

Moreover, the proposed method for constructing the different individual cohorts generates a cohort-specific heteroscedasticity that must also be modeled together with the existing cohort autocorrelation for observations occurring at different time points.

- For the variable year, we have defined four indicator variables, corresponding to each of the years 2006, 2010, 2014 and 2018, respectively. Thus, the year 2002, will be the reference level and the model analyzes the evolution of the average hourly wage of natives with respect to this reference year.

- For the gender variable, we define one indicator variable for males (m) and, consequently, females will be the reference level.

- For the education level variable, we have defined four indicator variables, corresponding to secondary education, professional training and higher education, respectively. In this way, primary level will be the reference level.

- For the working experience variable, four indicator variables were defined, corresponding to 11-20, 21-30, and more than 30 years of working experience, respectively. In this way, the group of 0-10 years of work experience will be the reference level.

- Finally, the immigration rate for each observation has been computed for each available pseudo panel. That is, as an example, for the group of males with high school education and 11-20 years of experience in 2010, the number of immigrants with these characteristics were counted and this number was then divided by the number people (natives and immigrants) having those characteristics. At the end, we will have the immigration rate for each pseudo panel. In this sense, a value of 0.1 means that from the number of males with high school education and 11-20 years of working experience in 2010, 10% were immigrant workers. In conclusion, the immigration rate is going to be a continuous variable.

In addition, we have also considered it relevant to include the corresponding two-way interaction terms for variables the categorical available here, so that, if required, the model can be better interpreted, and its results understood within the context of the specific available dataset.

Therefore, the proposed multiple linear regression model for the response variable given by the average hourly wage of native workers, *W*, for the cohort *i* in the time *t*, will be given by:

$$Log\ (w_{it}) = \beta_0 + \omega\ Rate_{it} + \sum_{j=1}^{4} \beta_j\ Year(j)_{it} + \lambda\ Male_{it} + \sum_{j=1}^{3} \gamma_j\ Education(j)_{it} + \sum_{j=1}^{3} \varphi_j\ Experience(j)_{it} + o\ (Rate_{it} *\ Male_{it}) + \sum_{j=1}^{4} \varsigma_j\ (Rate_{it} * Year(j)_{it}) + \sum_{j=1}^{4} \eta_j\ (Year(j)_{it} * Male_{it}) + \sum_{j=1}^{3} \psi_j\ (Rate_{it} * Education(j)_{it}) + \sum_{j=1}^{3} \theta_j\ (Male_{it} * Education(j)_{it}) + \sum_{j=1}^{3} \iota_j\ (Rate_{it} * Experience(j)_{it}) + \sum_{j=1}^{3} \vartheta_j\ (Male_{it} * Experience(j)_{it}) + \sum_{j=1}^{4} \sum_{k=1}^{3} \mu_{jk}\ (Year(j)_{it} * Experience(k)_{it}) + \sum_{j=1}^{4} \sum_{k=1}^{3} \xi_{jk}\ (Year(j)_{it} * Education(k)_{it}) + \sum_{j=1}^{3} \sum_{k=1}^{3} \zeta_{jk}\ (Education(j)_{it} * Experience(k)_{it}) + \varepsilon_{it}\ ,$$

$$i = 1, 2, \dots , 160;\ t = 0, 1, 2, 3, 4, \qquad (1)$$

where ε_{it} is the random error, assumed to follow a normal distribution with mean equal to zero, a non-constant variance, $Var(\varepsilon_{it})= \sigma^2{}_{it}$, and covariance different from zero for the same cohort, $Cov\ (\varepsilon_{it}, \varepsilon_{is}) = \sigma_{it,is}$,but equal to zero for different cohorts, $Cov\ (\varepsilon_{it}, \varepsilon_{js}) = 0, \forall_i \neq_j$.

In model (1), we would like to mention again that the reference levels for gender (male), year (2002), education (primary studies) and experience (0-10 years of working experience) are not included in the model, so the model constant, β_0 , can be interpreted as que average hourly wage for the reference cohort. This is, a male with primary studies, 0-10 years of working experience for the year 2002.

The model was estimated by ordinary least squares and standard deviations for the parameters in the model were calculated by applying the heteroscedasticity and autocorrelation corrections proposed and described in Arellano (1987), which allows a completely general structure with respect to heteroscedasticity and serial correlation. The model estimation and the corresponding standard deviation correction were implemented in the statistical software program R (R Core Team, 2014), using the plm function in the plm library.

6. Empirical results

Table 4 includes the results obtained from the fitting of model (1), including the estimates, standard deviations, as well as the statistical significance test statistic of the parameters associated with the different variables included in the model, and the corresponding p-values associated with these bilateral tests. For these contrasts, we have set a significance level of $\propto = 0.05$.

Table 4: Estimated parameters for the multiple linear regression model in equation (1), including the corresponding standard deviations, test statistics and p-values for the parameters included in the model.

	Estimate	Std. Error	t-value	Pr (>|t|)
($\hat{\beta}_0$) Const	0.75770	0.03568	21.23770	<0.0001
Rate (ω)				
($\hat{\omega}$) Immigration rate	0.00780	0.54114	0.01440	0.98853
Year (β_j)				
($\hat{\beta}_1$) 2006	0.03660	0.01669	2.19240	0.03080
($\hat{\beta}_2$) 2010	0.1234	0.022495	5.49	<0.0001
($\hat{\beta}_3$) 2014	0.12034	0.016432	7.32	<0.0001
($\hat{\beta}_4$) 2018	0.096401	0.021114	4.57	<0.0001
Gender (λ)				
($\hat{\lambda}$) Male	0.15472	0.0088172	17.5473	<0.0001
Education (γ_j)				
($\hat{\gamma}_1$) High School	0.039929	0.026209	1.52	0.13
($\hat{\gamma}_2$) Professional training	0.095951	0.033095	2.89920	0.00467
($\hat{\gamma}_3$) Higher studies	0.38703	0.029564	13.0916	<0.0001
Experience (φ_j)				
($\hat{\varphi}_1$) 11-20 years	0.16755	0.035326	4.743	<0.0001
($\hat{\varphi}_2$) 21-30 years	0.23243	0.036739	6.3266	<0.0001
($\hat{\varphi}_3$) > 30 years	0.30851	0.036699	8.4065	<0.0001
Rate * Gender (o)				
($\hat{o}$) Rate * Male	-0.24087	0.049367	-4.8793	<0.0001
Rate * Year (ς_j)				
($\hat{\varsigma}_1$) Rate * 2006	0.16483	0.20417	0.8073	0.4215588
($\hat{\varsigma}_2$) Rate * 2010	0.20058	0.21991	0.9121	0.36408
($\hat{\varsigma}_3$) Rate * 2014	0.13076	0.19807	0.6602	0.5107851
($\hat{\varsigma}_4$) Rate * 2018	0.2706	0.28489	0.9498	0.344657
Year * Gender (η_j)				
($\hat{\eta}_1$) 2006 * Male	0.0026377	0.0057176	0.4613	0.6456315
($\hat{\eta}_2$) 2010 * Male	-0.00935	0.0057205	-1.6344	0.1055488
($\hat{\eta}_3$) 2014 * Male	-0.015643	0.0055752	-2.8059	0.0061116
($\hat{\eta}_4$) 2018 * Male	-0.002047	0.0059029	-0.3468	0.7294988

Rate * Education (ψ_j)				
($\hat{\psi}_1$) Rate * High school	0.16875	0.14178	1.1903	0.2369768
($\hat{\psi}_2$) Rate * Professional training	0.23858	0.43424	0.5494	0.5840358
($\hat{\psi}_3$) Rate * Higher studies	0.9517	0.43424	-4.4008	<0.0001
Gender * Education (θ_j)				
($\hat{\theta}_1$) Male $*$ High school	-0.023783	0.0052125	-4.5627	0.0000154
($\hat{\theta}_2$) Male $*$ Professional training	-0.0095379	0.0055676	-1.7131	0.0900213
($\hat{\theta}_3$) Male $*$ Higher studies	0.010209	0.0054339	1.8787	0.0634125
Rate * Experience (ι_j)				
($\hat{\iota}_1$) Rate * 11-20	0.67345	0.3056	2.2037	0.0300149
($\hat{\iota}_2$) Rate * 21-30	0.73048	1.1952	0.6112	0.5425785
($\hat{\iota}_3$) Rate * >30	2.978	1.3317	2.2363	0.0277256
Gender * Experience (ϑ_j)				
($\hat{\vartheta}_1$) Male*11-20	-0.0046768	0.0071466	-0.6544	0.5144602
($\hat{\vartheta}_2$) Male*21-30	-0.0089007	0.0076256	-1.1672	0.2461065
($\hat{\vartheta}_3$) Male*>30	-0.013552	0.0077268	-1.7539	0.082747
Year * Experience (μ_{jk})				
($\hat{\mu}_{11}$) 2006*11-20	-0.006479	0.014824	-0.4371	0.663086
($\hat{\mu}_{21}$) 2010*11-20	-0.039211	0.020649	-1.8989	0.0606788
($\hat{\mu}_{31}$) 2014*11-20	-0.064792	0.01374	-4.7154	<0.0001
($\hat{\mu}_{41}$) 2018*11-20	-0.055594	0.017416	-3.1922	0.001967
($\hat{\mu}_{12}$) 2006*21-30	-0.011783	-0.015462	-0.7621	0.4479497
($\hat{\mu}_{22}$) 2010*21-30	-0.018544	0.021635	-0.8571	0.3936659
($\hat{\mu}_{32}$) 2014*21-30	-0.038135	0.015028	-2.5375	0.0128275
($\hat{\mu}_{42}$) 2018*21-30	-0.01087	0.019718	-0.5166	0.6066482
($\hat{\mu}_{13}$) 2006*>30	-0.018836	0.015592	-1.2081	0.2300705
($\hat{\mu}_{23}$) 2010*>30	-0.023176	0.021719	-1.0671	0.2886842
($\hat{\mu}_{33}$) 2014*>30	-0.037511	0.015405	-2.4350	0.0167982
($\hat{\mu}_{43}$) 2018*>30	-0.0098013	0.019848	-0.4938	0.6226061

Year * Education (ξ_{jk})				
($\hat{\xi}_{11}$) 2006*High School	0.0043577	0.0084599	0.5151	0.6077
($\hat{\xi}_{21}$) 2010*High School	0.0054872	0.0085383	0.6427	0.522027
($\hat{\xi}_{31}$) 2014*High School	0.011629	0.0084022	1.384	0.169669
($\hat{\xi}_{41}$) 2018*High School	0.012296	0.0091548	1.3432	0.1824833
($\hat{\xi}_{12}$) 2006*Professional training	-0.0006169	0.0088464	-0.0697	0.9445574
($\hat{\xi}_{22}$) 2010*Professional training	-0.000081	0.0087918	-0.0092	0.992668
($\hat{\xi}_{32}$) 2014*Professional training	0.03142	0.0088197	3.5625	0.0005817
($\hat{\xi}_{42}$) 2018*Professional training	0.055733	0.01008	5.5289	<0.0001
($\hat{\xi}_{13}$) 2006*Higher studies	-0.0045234	0.0087049	-0.5196	0.6045501
($\hat{\xi}_{23}$) 2010*Higher studies	0.025558	0.0086304	2.9614	0.003887
($\hat{\xi}_{33}$) 2014*Higher studies	0.017239	0.0087969	1.9597	0.0530304
($\hat{\xi}_{43}$) 2018*Higher studies	0.053853	0.0096482	5.5816	<0.0001
Education * Experience (ζ_{jk})				
($\hat{\zeta}_{11}$) High school * 11-20	0.046098	0.026446	1.7431	0.0846149
($\hat{\zeta}_{21}$) Professional training* 11-20	0.040048	0.032808	1.2207	0.2252922
($\hat{\zeta}_{31}$) Higher studies* 11-20	-0.025966	0.028781	-0.9022	0.3692844
($\hat{\zeta}_{12}$) High school * 21-30	0.068453	0.027693	2.4718	0.015261
($\hat{\zeta}_{22}$) Professional training* 21-30	0.050335	0.03358	1.499	0.1372694
($\hat{\zeta}_{32}$) Higher studies* 21-30	-0.054552	0.03115	-1.7513	0.0831919
($\hat{\zeta}_{13}$) High school * >30	0.049106	0.02783	1.7645	0.080931
($\hat{\zeta}_{23}$) Professional training* >30	0.016871	0.033769	0.4996	0.6185331
($\hat{\zeta}_{33}$) Higher studies * >30	-0.11802	0.032372	-3.6457	0.0004392

As for the goodness-of-fit for the proposed model, the coefficient of determination R^2=0.998, which means that 99.8% of the variability of the response variable is explained by the proposed model and, therefore, it represents a good model proposal and it also provides a good model fit for the dataset under study.

At the $\propto$= 0.05 significance level, the effect of the Immigration Rate is not statistically significant, whereas the effects of the variables Year, Gender, Education level and Experience are statistically significant, except for the High school education level.

With regard to the variable Year and taking 2002 as the reference year, we observe that, in 2006, the logarithm of the average hourly wage of native workers in Spain was 0.037 and then 0.123 in 2010, which means that there was an increasing trend in the logarithm of the average hourly wage of native workers until 2010. However, in 2014 there is a very small decrease of 0.003 points in our logarithm scale with respect to 2010, and, finally, a slightly larger decrease in the year 2018 (0.024 in terms of the logarithm of the average hourly wage with respect to the previous year). As for the Gender variable, we can see that, in general, male native workers earn 0.155 more than females, always in terms of the logarithm of the average hourly salary. Consequently, there is a statistically significant wage gap between males and females. Regarding the Education and Experience variables, we can observe that, for higher education levels and more years of experience, the logarithm of the average hourly salary increases and that, except for the high school education level, this increase is statistically significant for all the different categories for these variables.

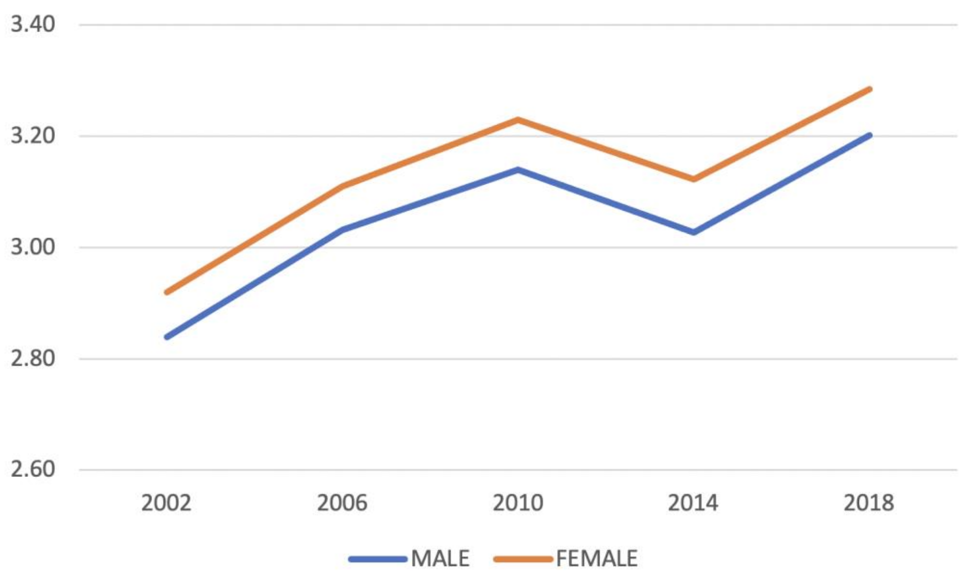
With regard to the interaction terms included in the model, the interaction between the Rate and Gender is statistically significant and, in addition, it is negative for the male category, suggesting that, regardless of the Immigration rate value, the estimated for the logarithm of the average hourly wage will be smaller for males than for females. However, no interaction between the Rate and Year variables are statistically significant. As for the Year and Gender variables, it is only statistically significant for the year 2014, which means that males behave differently for this year. This behavior can be clearly verified in the estimated coefficients (all negative ones except for 2006) for the regression model for these two interaction terms (see Table 4), with the coefficient for 2014 being much larger, in absolute value, to those for the remaining years. Moreover, there is only one significant category in terms of the interaction between the Rate and Education variables, which corresponds to the Higher studies education level. We can observe (see Table 4) the increasing trend in the estimated positive coefficients for this interaction term, suggesting that the higher the education level, in average, the higher the estimated of the logarithm of the average hourly salary. In addition, for the interaction between the Gender and Education variables, it is significant only for the high school studies education level. From the information reported in Table 4, we can see that the High school and Professional training education levels have a negative estimated coefficient, whereas the Higher studies category has a positive estimated coefficient.

As for the interaction between the Rate and Experience variables, it is only statistically significant for the 11-20 and >30 years of experience categories. Moreover, from the estimates reported in Table 4, we can observe that more years of experience, in average, results in a higher estimate for the logarithm of the average hourly salary. With regard to the interaction between the Gender and Experience variables, none of the categories are statistically significant.

Additionally, for the interaction between the Year and Experience variables, the only significant effects are those for the 2014 and 11-20, 2018 and 11-20, 2014 and 21-30 and 2014 and >30 categories. Therefore, it seems that, regardless of the years of working experience, this term is significant for the year 2014 and it is not significant for the year 2006, whereas for the remaining years, it is only significant for the 2018 and 11-20 category. As for the interaction between the Year and Education variables, the only significant effects are those for 2014 and Professional training, 2018 and Professional Training , 2010 and Higher studies and 2018 and Higher studies. Finally, for the interaction between the Education and Experience variables, the only significant effects are those for the High school and 21-30 and Higher studies and >30 categories.

In order to be able to observe the effect of the variables Gender and Year on the average hourly wage, Figure 37 includes, in a logarithm scale, their average hourly wage for the Higher studies education and 11-20 years of working experience categories. As an aside, we would like to mention that, in order to generate these figures, the Immigration rate was taken into account in the model estimates for equation (1).

Figure 37: Evolution of the logarithm of the average hourly wage for female and male with higher studies education and 11-20 years of experience, taking 2002 as the reference year for each gender.



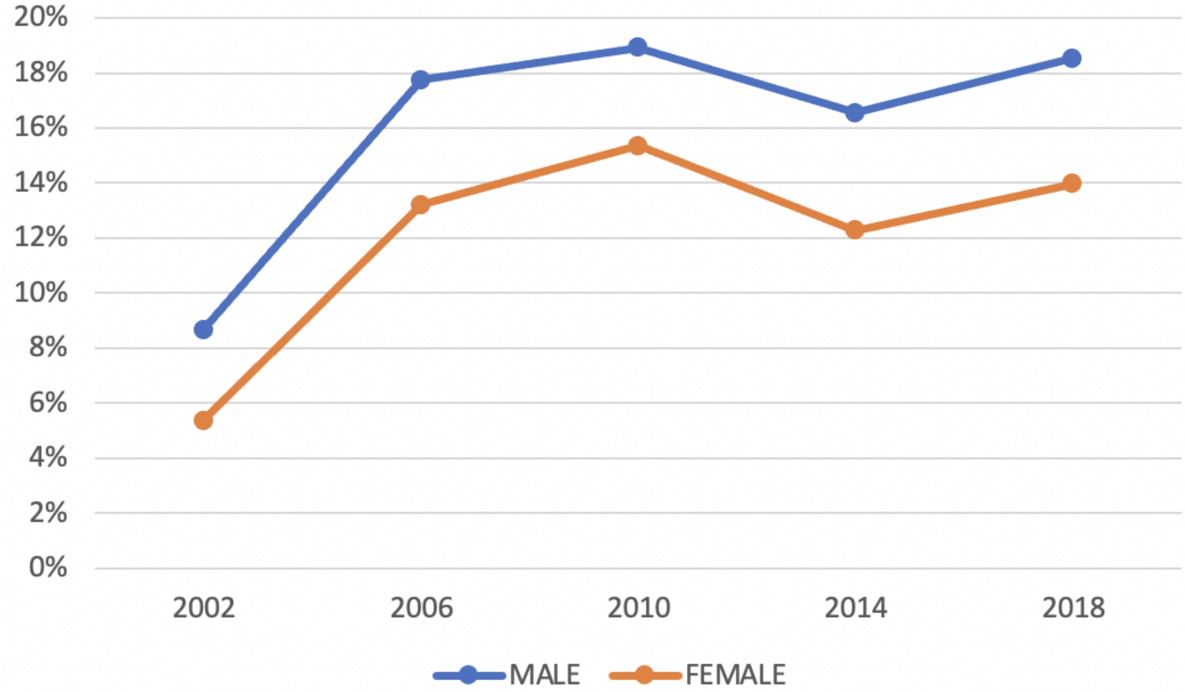
Source: Own elaboration based on the model estimates in Table 4.

As can be observed in Figure 37, and keeping in mind that estimated coefficient for the Immigration Rate and Gender interaction was negative and significant for the male category, the estimated logarithm of the average hourly wage for females is higher than that of males. In Figure 37, we can observe an increase until 2010 (logarithm of the average hourly wage of 3.14 for males and 3.23 for females). However, there is a significant fall in 2014 (a decrease of 0.113 points for males and 0.106 for females). Finally, in 2018, there is a new rise but, in this specific year, it is higher than the one in 2010. More specifically, for 2018, male workers have a logarithm of the average hourly wage equal to 3.201, whereas that for females was equal to 3.28.

Finally, we have already mentioned that the effect of the immigration rate was the only variable not being statistically significant. Consequently, our immigration rate variable, computed for each available observation/pseudo panel, is the percentage of immigration in each group. However, it may also be worth mentioning than only few interactions between Rate and the other variables were statistically significant. In addition, we have observed that the negative estimate obtained in for the Rate and Gender (i.e. male) interaction resulted in a lower hourly average salary for male workers when compared to female workers. Moreover, the non-significant result obtained for the effect of the Rate variable may be interpreted in a way such that an increase or a decrease of this rate does not make any significant difference on the average hourly salary of natives workers (without taking Gender into account). Nevertheless, the interaction between Gender and Rate is statistically significant and has a considerable negative estimate. This means that immigration has a negative impact on the hourly average salaries for male native workers.

As for the immigration rate in the 160 observations, this rate was very low, not even 10%, which means that in the different groups there were not even 10 immigrants out of every 100 people. Figure 38 shows a graph with the evolution of immigration rates for males and females with primary education and 0-10 years of working experience. In conclusion, male immigration rates have always been higher in comparison to that of female workers.

Figure 38: Temporal evolution of the immigration rate for male and female workers with primary education and 0-10 years of experience.



Source: Own elaboration based on the model estimates reported in Table 4.

7. Conclusion

In this work, we have analyzed the different factors that influence the average hourly salary of natives in Spain. In order to be able to draw some conclusions with regard to this matter, we have used a multiple linear regression model approach. The data used for this work was obtained from the Spanish National Institute of Statistics (INE). Our methodological approach was based on a pseudo-panel of correlated data that have a specific cohort heteroscedasticity behavior, and we have proposed a very general and flexible multiple linear regression model in order to model this type of data.

The results obtained in this study have shown the clear relationship between the available variables and the average hourly wage. Gender, year, education level and working experience have been identified as significant contributing factors affecting the average wage of natives in Spain.

We have observed that those having the highest average hourly wage are the workers with the highest level of education and working experience. Moreover, in general, male workers have a higher salary than females during the entire period under study. However, we have also concluded that the effect of the immigration rate variable is not statistically significant, although some of the interactions between the remaining variables and rate turned out to be significant. Consequently, the average hourly wage of the native male decreases due to the effect of the interaction between the Gender and Rate variables. Finally, we have been able to verify that the average hourly salary increases until 2010, where it reaches a maximum and, then, there is a decrease until 2014, which gradually recovers for the year 2018.

We can also add that during the years under study, from 2002 to 2018, males earned more than female workers in Spain. This leads us to the general conclusion that one of the most important determinants of the average hourly salary in Spain is the Gender of the worker. Moreover, we also concluded that workers having more experience and a higher education level would be the ones having higher average hourly wages. However, it is clear that these two factors largely depend on the worker him/herself, which does not occur for the case of the Gender variable.

Therefore, as a final conclusion of this study, we observe that, in relation to the average hourly realized salary, there are differences in gender, education level and working experience of native workers, as well as on the different years in the period under study.

As we have already mentioned at the beginning of this work, the vast majority of foreigners in Spain come from European countries, followed by Morocco and Colombia. The effect of immigration has indeed turned out to be important in determining the average hourly wage of native workers in Spain. However, something that cannot be verified, because the available database of the Spanish National Institute of Statistics used for this analysis does not determine from which country the foreigner comes from, is whether the specific origin of the foreign worker influences the average hourly wage in Spain. That is to say, the dataset only provides information on whether the worker is Spanish or foreigner.

To conclude, we believe it is essential that, when setting salaries for their workers, companies and public institution take these results into consideration. If measures are taken to address the wage gap between men and women, the existing differences could decrease. This would be a big step into the future. However, it is true that many companies have the same salaries for men and women. Despite this big recent progress, many of the most important positions in companies are occupied by male workers. Consequently, we propose that the Spanish National Institute of Statistics should improve the labelling used for the variable describing the working positions in the microdata used for this study, mainly because it included odd working positions titles, which were found to be unrealistic and, therefore, because of this, we decided not to take them into consideration for the study.

8. Bibliography

Arellano, M. (1987). Computing robust standard errors for within group estimators. *Oxford Bulletin of Economics and Statistics*, **49**, 431-434.

Atienza Montero, P. y Romo Calixto, C.E. (2020). El efecto de la población inmigrante sobre el empleo y salarios de los trabajadores nativos. Una investigación empírica para España, 2009-2018. *Estudios de Economia Aplicada*, **39(3),** 1-24.

Borjas, G.J. (2003). The labor demand curve id downward sloping: reexamining the impact of immigrants on the labor market. *The Quarterly Journal of Economics*, **118(4)**, 1335-1374.

Carrasco, R., Jimeno, J.F. y Ortega, C. (2008). The effect of immigration on the labor market performance of native-born workers: some evidence for Spain. *Journal of Population Economics*, **21(3)**, 627-648.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, **30**, 109–126.

EUROSTAT (2022). Database: Migration and migrant population statistics. Available in: https://ec.europa.eu/eurostat/en/web/products-datasets/-/MIGR_POP3CTB

Gerfin, M. y Kiaser, B. (2010). The effects of immigration on wages: an application of the structural skill-cell approach. *Swiss Society of Economics and Statistics*, **146(4),** 709-739.

González, A. (2002). Efectos macroeconómicos de la inmigración, impacto sobre el empleo y los salarios de los nativos. *Papers. Revista de Sociologia*, **66**, 133-153.

Jaumotte, F., Koloskova, K., Saxena, S. (2016). Migrants bring economic benefits for advanced economies. *IMF*. https://www.imf.org/en/Blogs/Articles/2016/10/24/migrants-bring-economic-benefits-for-advanced-economies

INE (2001). Database: Population Census. Available in: https://www.ine.es/censo/es/inicio.jsp

INE (2002). Database: Quadrennial Salary Structure Survey. Available in: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177025&menu=resultados&secc=1254736061996&idp=1254735976596#!tabs-1254736195110

INE (2006). Database: Quadrennial Salary Structure Survey. Available in: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177025&menu=resultados&secc=1254736061996&idp=1254735976596#!tabs-1254736195110

INE (2010). Database: Quadrennial Salary Structure Survey. Available in: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177025&menu=resultados&secc=1254736061996&idp=1254735976596#!tabs-1254736195110

INE (2011). Database: Population Census. Available in: https://www.ine.es/censos2011/tablas/Wizard.do?WIZARD=1&reqCode=paso1

INE (2014). Database: Quadrennial Salary Structure Survey. Available in: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177025&menu=resultados&secc=1254736061996&idp=1254735976596#!tabs-1254736195110

INE (2018). Database: Quadrennial Salary Structure Survey. Available in: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177025&menu=resultados&secc=1254736061996&idp=1254735976596#!tabs-1254736195110

INE (2020a). Database: Salary Structure Survey. Available in: https://www.ine.es/prensa/ees_2020.pdf

INE (2020b). Database: Wages, income, social cohesion. Available in: https://www.ine.es/jaxiT3/Tabla.htm?t=10885

INE (2021). Database: Population Census. Available in: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176992&menu=resultados&idp=1254735572981#!tabs-1254736195781

INE (2022). Database: Foreign Population by Nationality, Communities, Gender and Year. Available in: https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=02005.px&L=0

Maddala, G. S. (2001). *Introduction to Econometrics*. Macmillan Publishing Company: New York.

R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.