



UNIVERSIDAD
COMPLUTENSE
MADRID

FACULTAD DE CIENCIAS ECONÓMICAS
Y EMPRESARIALES

GRADO EN ECONOMÍA
TRABAJO DE FIN DE GRADO

TÍTULO: *What drives labor migration?*

AUTOR: *Daniel Pérez García*

TUTOR: *Juan César Palomino*

CURSO ACADÉMICO: *2023-2024*

CONVOCATORIA: *Junio*

Abstract

The goal of this paper is to measure what are the main driving forces behind internal migration using regional data from Spain. Some of the explanatory variables that will be used are: the average wage of each region, renting costs or unemployment rate among others. The objective is to see the effect and significance of these variables in explaining both the entrance and exit of internal labor migrants for the Spanish regions. I will also try to analyse this effects separating migrants by their educational level and economic sector. Results show that wages, unemployment, and renting costs are significant variables for the overall regional labor mobility, but these driving forces vary by sector and education level.

Key words: Labor Migration, Labor Market, Regional Inequalities.

INDEX

1	Introduction	3
2	Literature Review	4
3	Theoretical Framework	6
3.1	Entry and exit functions	8
4	Data	11
5	Descriptive analysis	13
5.1	Total labor mobility	13
5.2	By economic sector	16
5.3	By educational level	20
6	Methodology	23
6.1	Regional control	23
6.2	Temporal control	24
6.3	Estimated equations	24
7	Empirical Findings	27
7.1	By economic sector	31
7.2	By educational level	37
8	Conclusions	42
	References	44
	Appendix	46

1 Introduction

Regional inequalities in Spain are present in almost any indicator that we look at, and they have persisted throughout the years. But, undoubtedly, these inequalities are intensified when looking at the labor market. The huge differences regarding both remuneration of workers and working opportunities make that thousands of workers migrate each year in order to find a job or enhance their working conditions. Therefore, trying to analyze what are the driving forces behind this decisions could help us to delve into the challenging issue that regional inequality represents.

After reviewing what has been done previously in the field of labor migration, I will try to construct a theoretical model trying to explain what factors affect both the decision to migrate and the entrance of workers into a region. I have self-elaborated an extensive database including different variables that reflect the labor and economic situation of the Spanish regions for the period 2006-2022. In addition to the variables indicating regional labor mobility, I have included other relevant indicators such as the average wage, renting costs, unemployment rate or industrial and educational structure among others.

Using this database I will test if my theoretical conjectures have empirical support and explain the results obtained from my econometric analysis. Finally, I will test these models separating migrants by educational level and working sector.

What we appreciate is that wages, unemployment and renting costs are determinant factors for the overall regional mobility rates. However, when considering different economic sectors and educational levels, the driving forces vary according to the specific characteristics of each group.

The structure of this work is as follows: In [Section 2](#) I will review the

literature regarding labor mobility, in [Section 3](#) I will introduce the theoretical framework on which my empirical analysis will be based, in [Section 4](#) I will describe the dataset that I will be using. [Section 5](#) presents some descriptive analysis and [Section 6](#) the models that will be estimated. [Section 7](#) presents the empirical results and [Section 8](#) concludes.

2 Literature Review

The main theoretical framework trying to explain the reason of labor mobility is known as the *Human Capital model of migration*. It argues that the decision of a worker on whether to migrate or not lies on the comparison between the net discounted returns of migrating and the expected returns of staying. Apart from the monetary costs of labor mobility, uncertainty should also be taken into account ([Kan, 1999](#); [Bentolila, 2001](#); [Khwaja, 2002](#)) In addition to this monetary costs there exists non-monetary costs related to migration ([Sjaastad, 1962](#)).

This theoretical framework has been expanded by the introduction of other factors apart from the monetary ones. There was evidence found that improving quality of life, even though reducing consumption is also a driving force for migrates specially moving from large cities ([Hamilton and Mail, 2003](#); [Chhetri et al., 2009](#)).

Apart from this theoretical aspect, there are other classical arguments that explain the phenomena of labor migration as the desire to secure employment instead of seeking income maximization ([Simon, 1957](#); [Borjas et al., 1992](#)).

Most of the empirical works regarding labor mobility have gone in the line of analyzing the consequences of labor mobility in both the person who migrates and the region receiving migration.

In the field of analyzing the short and medium term effects of labor migration on migrants' wages, [Roca and Puga \(2017\)](#) work deserves to be highlighted, as the authors show (using Spanish data) that the workers that migrate to bigger cities still earn a higher wage when they reallocate to a smaller city.

[Pronchev et al. \(2019\)](#) and [Hamidov \(2022\)](#) analyze foreign migration, using as explanatory causes the economic situation of the country of origin. The authors note some consequences for both the migrant and also for the destination country, pointing out that they are commonly negative for the former and positive for the latter.

However, there is little work on the field of analysing the effects that drive labor migration. [Morrison and Clark \(2011\)](#) look at the micro motives of overall migrations in New Zealand, trying to align its findings with the theoretical bases of [Simon \(1957\)](#) and [Borjas et al. \(1992\)](#).

Regarding the literature dealing with Spanish evidences on labor migration [Bentolila \(2001\)](#) and [Liu \(2018\)](#) stress the fact that labor migration in Spain is much lower compared to other European countries. The main empirical findings are the following:

Using a semi-parametric analysis, [Maza and Villaverde \(2004\)](#) found that the main variables explaining inter regional net migrations in Spain in the last 90s were differentials in GDP per capita and climate conditions, followed by differentials in unemployment rates and housing costs.

[Bover and Arellano \(2002\)](#) estimated a multinomial model of the probability of intra-regional migration by town size of origin and destination and found statistical significance in explanatory variables such as housing price, the share of industry workers, the unemployment rate or the educational level.

[Llorente Jiménez \(2020\)](#) tries to explain the absolute number of labor

migrations in Spain by running a regression against GDP, temporal contracts and permanent contracts. However, the article only analyses overall regional migration at a national level.

Finally, other works that deserve to be mentioned and that use econometric analysis to study regional migration in Spain have been done by [Angulo and Mur \(2005\)](#); [Mulhern and Watson \(2010\)](#) and [Clemente et al. \(2016\)](#).

3 Theoretical Framework

The idea of this section is to develop a theoretical model which compiles relevant economic variables that could affect the entry and exit of labor migrants within a region and try to argue which effect we expect them to have (positive or negative). This model will be micro based, we will center on individuals' decisions to explain migrations at a macro (regional) level.

For simplification I will analyze the case of a worker from region A that thinks of moving to region B. Based on the *Human Capital Model of Migration*, where the discounted returns of migrating are compared to the expected returns of staying we could define both concepts as following:

$$E(migrating_B) = P(job_B) \times E(wage_B) - migrating\ costs \quad (1)$$

$$E(staying_A) = P(job_A) \times E(wage_A) - staying\ costs \quad (2)$$

Let's analyze first the expected returns of migrating as expressed in (1).

- $P(job_B)$: It is the probability of finding a job in region B. It could be approximated by $1 - unemployment_B$, but it is also affected by other economic components of the region, such as the weight of that region in the whole economy, the economic structure or the economic institutions (that may have power to implement regional-specific labor

market policies). The economic situation of the country as a whole also alters this probability.

- $E(wage_B)$: We will assume that the expected wage is the average wage.
- *migrating costs*: It is important to distinguish 2 types of costs:
 - Monetary costs: Renting costs (We will assume no differences in main utilities among different regions).
 - Non-monetary costs: Costs associated to moving away from relative or friends are examples of these non-monetary costs. Geographical distance of migration could serve as an indicator for the magnitude of these costs (the closer the migration the lower the non-monetary costs). Given that in Spain there are 3 official languages apart from Spanish (Catalan, Galician and Basque), there could also be a cost derived from learning a new language when) moving to a region where any of these languages is spoken.

For the expected returns of staying, the probability of finding a job ($P(job_A)$) and the expected wage ($E(wage_A)$) are interpreted in the same way as if you migrate but for region A instead of region B.

However, the interpretation of the costs of staying differ, as we only consider the monetary costs (renting costs).

Thus, an individual will migrate if the following condition is met:

$$E(migrating_B) \geq E(staying_A) \tag{3}$$

Now that we know what is taken into account when deciding whether to migrate or not, we can build one function relating the driving forces of labor migration (Exit function) and another for labor immigration (Entry function).

3.1 Entry and exit functions

$$Entry_{i,t} = f(W_{i,t-1}, u_{i,t-1}, Rent_{i,t-1}, Economic\ structure_{i,t}, Geographical\ component_i) \quad (4)$$

(+)
(-)
(-)
(+)
(?)

$$Exit_{i,t} = f(W_{i,t-1}, u_{i,t-1}, Rent_{i,t-1}, Economic\ structure_{i,t}, Geographical\ component_i) \quad (5)$$

(-)
(+)
(?)
(-)
(?)

$Entry_{i,t}$ refers to the entry rate of region i in year t and $Exit_{i,t}$ is the exit rate of region i in year t . For these ratios I will use the following definition proposed by SEPE (Servicio Público de Empleo Estatal):

$$Entry\ rate_{i,t} = \frac{Entries_{i,t}}{Stays_{i,t} + Entries_{i,t}} \quad Exit\ rate_{i,t} = \frac{Exits_{i,t}}{Stays_{i,t} + Exits_{i,t}}$$

Where $Entries_{i,t}$ and $Exits_{i,t}$ stand for the number of workers that enter in and exit from region i in year t and $Stays_{i,t}$ refers to the number of contracts that remain from period $t - 1$ in region i at period t .

When looking at the entry and exit rates for the different economic sectors and educational levels I will use the following metric:

$$Entry\ rate_{i,j,t} = \frac{Entries_{i,j,t}}{Stays_{i,j,t} + Entries_{i,j,t}} \quad Exit\ rate_{i,j,t} = \frac{Exits_{i,j,t}}{Stays_{i,j,t} + Exits_{i,j,t}}$$

The interpretation is almost identical as before, but now including the sub index j that stands for the j th economic sector or educational level.

I use a 1-year lag for two reasons. First, labor mobility is usually a decision deliberated over a period of time. If an individual decides whether to move or stay, this decision was most likely made at least one period earlier. Second, individuals do not have real-time information for almost any economic variable such as wages or unemployment rate, so the reference value for deciding will most probably be the one referring to the previous year.

Theoretically, we should not expect any component appearing in one function but not in the other, as the decision to migrate takes into account the

situation of both the region of origin and of destination. After presenting equations (4) and (5) I will define the terms inside them:

- $W_{i,t-1}$: It is the expected wage of region i in $t-1$.
- $u_{i,t-1}$: It is the unemployment rate of region i in $t-1$.
- *Economic structure* $_{i,t}$: It is the economic structure of region i in $t-1$. Inside this variable there are comprised factors such as the distribution of the labor force within the different economic sectors (classified as agriculture, construction, industry and service sector), the educational level of the population, the share of temporary contracts or the GDP per capita growth of the region among others.
- $Rent_{i,t-1}$: it is the average renting cost (of an apartment) in region i in period $t-1$.
- *Geographical component* $_i$ It is a region specific, time invariant component that includes factors such as the lifestyle of the region, geographical factors (if it is near the sea, its weather. if it is outside the peninsula...), cultural components or the fact of leaving friends and family in your origin region (Maza and Villaverde, 2004). It is very important to take into account due to the differences between Spanish regions.

Lastly, I want to point out the symbols between parenthesis below each variable in both functions. This signs reflect the result I expect this variables to have when carrying out the empirical analysis.

Entry Function: For the *Entry Function* the interpretation is the following; The higher the average wage in region i last period, the more people will be willing to migrate to that region (higher expected net returns) and the same

happens with the economic structure of the region. The more developed this structure is, for example, the better the education or the better the sectorial structure in the region, the more workers will be willing to migrate there. However, a higher unemployment rate would imply a lower probability of finding a job there, and higher renting costs would imply lower net expected returns of migrating, thus, the lower the migration to region i will be. For the *geographical component* as it is region-specific a single sign cannot be attributed to it. It will have a positive effect on the entry rate for the regions with better qualitative conditions such as weather or, if we were looking at an individual level, that have friends or relatives living there and negative sign in the contrary scenario.

Exit Function: For the *Exit Function* the signs are reversed (with the exception of the renting costs, that will be expounded in [Section 6.3](#)) and the explanation is the same as in the *Entry Function*; regions with higher wages and more developed economic institutions are expected to have lower emigration rates, whereas regions with higher unemployment rates and higher living costs are thought to have higher exit rates. The explanation for the geographical component is the same as before, but changing the explanation of positive effect with negative effect and vice versa.

4 Data

My idea has been to construct a large dataset for all Spanish autonomous communities for the period 2006-2022, where, apart from data regarding labor mobility, I could also have data regarding their economic and social situation. The sources where I have obtained these variables are:

- **Servicio Público de Empleo Estatal (SEPE):** From its *Observatorio de las Ocupaciones* I have taken all the data regarding labor mobility for all the regions in Spain. This data not only contains information at the aggregate level of the region, but also information dis-aggregated by sex, age, educational level or occupational sector ([Appendix A](#) compares it with other available datasets for labor migration in Spain)
- **Instituto Nacional de Estadística (INE):** All the information regarding demography of the region, educational level of the region, or the GDP per capita of the different regions was taken from here. Furthermore, several information was taken from two of the most significant surveys of the INE:
 - *Encuesta de Población Activa (EPA):* All the information regarding labor market, such as the distribution of wages, the average wage, the number of people employed in different sectors or the unemployment or employment rates were obtained throughout the *EPA*.
 - *Encuesta de Condiciones de Vida (ECV):* I have used its micro-data to obtain income data by educational level and economic sector for the different regions ([Appendix D](#) delves into this process).

- **Idealista:** Idealista is a Spanish company which operates as a Real Estate portal and that provides data regarding the square meter (m^2) price of renting houses in the Spanish regions.

Before moving on to describing the data using graphical analysis, [Figure 1](#) shows a correlation matrix for the main variables that I will be using in my empirical analysis.

Figure 1: Correlation matrix of main variables

	Entry rate	Exit rate	Average wage	Unemployment rate	Average rent (m^2)	Gini Index	Temporary jobs	Occupied with higher education	Non-tourism services	Occupied in agriculture
Entry rate	1	0.78	0.25	-0.18	-0.02	0.05	-0.34	0.09	-0.21	-0.03
Exit rate	0.78	1	0.02	-0.05	-0.4	-0.08	-0.22	-0.04	-0.37	0.24
Average wage	0.25	0.02	1	-0.31	0.55	0.27	-0.75	0.9	0.23	-0.53
Unemployment rate	-0.18	-0.05	-0.31	1	-0.47	0.55	0.21	-0.28	0.43	0.3
Average rent (m^2)	-0.02	-0.4	0.55	-0.47	1	-0.11	-0.36	0.41	0.36	-0.69
Gini Index	0.05	-0.08	0.27	0.55	-0.11	1	-0.34	0.28	0.44	-0.15
Temporary jobs	-0.34	-0.22	-0.75	0.21	-0.36	-0.34	1	-0.65	-0.13	0.58
Occupied with higher education	0.09	-0.04	0.9	-0.28	0.41	0.28	-0.65	1	0.2	-0.46
Non-tourism services	-0.21	-0.37	0.23	0.43	0.36	0.44	-0.13	0.2	1	-0.24
Occupied in agriculture	-0.03	0.24	-0.53	0.3	-0.69	-0.15	0.58	-0.46	-0.24	1

Source: Self-elaborated from SEPE and INE data.

5 Descriptive analysis

The idea of this section is to provide some illustrative information in order to know the basic characteristics of labor mobility in Spain. I will divide this part as follow; firstly I will look at total labor mobility (both entry rates and exit rates) and, afterwards, I will look at labor mobility separating by economic sector and by educational level.

It is very important to note that the data for the different economic sectors and educational levels only goes from 2016 to 2022, while the total rates for the regions go from 2006 to 2022.

5.1 Total labor mobility

Table 1: Summary statistics total of labor mobility data

	mean	sd	min	max
Entry Rate	0.1091	0.0567	0.0281	0.2662
Exit Rate	0.1176	0.0562	0.0402	0.3018
Mobility Rate	0.1138	0.0534	0.0372	0.2836
Observations	289			

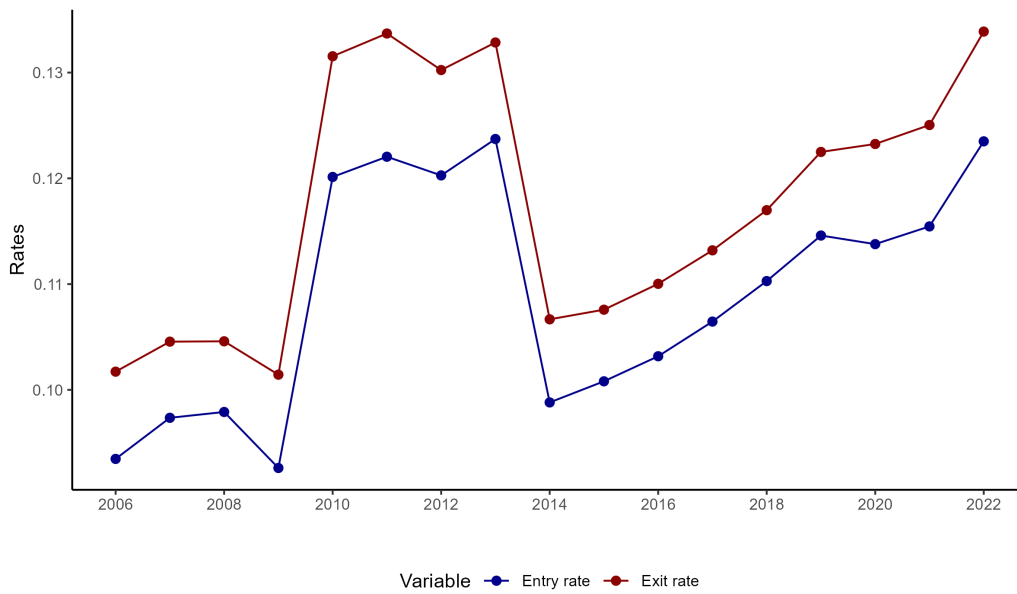
Note: Self-elaborated from SEPE data.

[Table 1](#) shows the main statistics from the entry rates, exit rates and mobility rates from 2006 to 2022. As we appreciate, the exit rates exhibit (on average) higher values than the entry rates. However, the standard deviations for both metrics are almost identical.

[Figure 2](#) shows the time evolution of the average entry and exit rates in

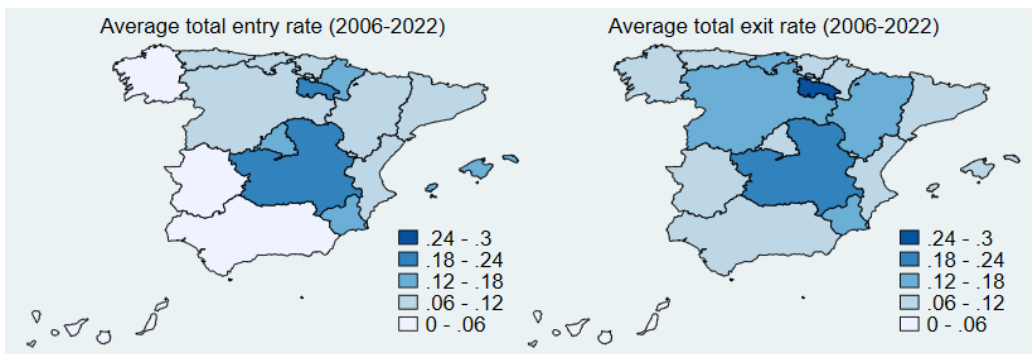
Spain for the period 2006 to 2022. As we appreciate, there was a huge increase during the years 2010 to 2013 due to the global financial crisis that dramatically affected the Spanish labor market. During the last years we observe an increasing tendency from both entry and exit rates.

Figure 2: Average entry and exit rates by year (2006-2022).



Source: Self-elaborated from SEPE data.

Figure 3: Average entry and exit rates by region (2006-2022).



Source: Self-elaborated from SEPE data.

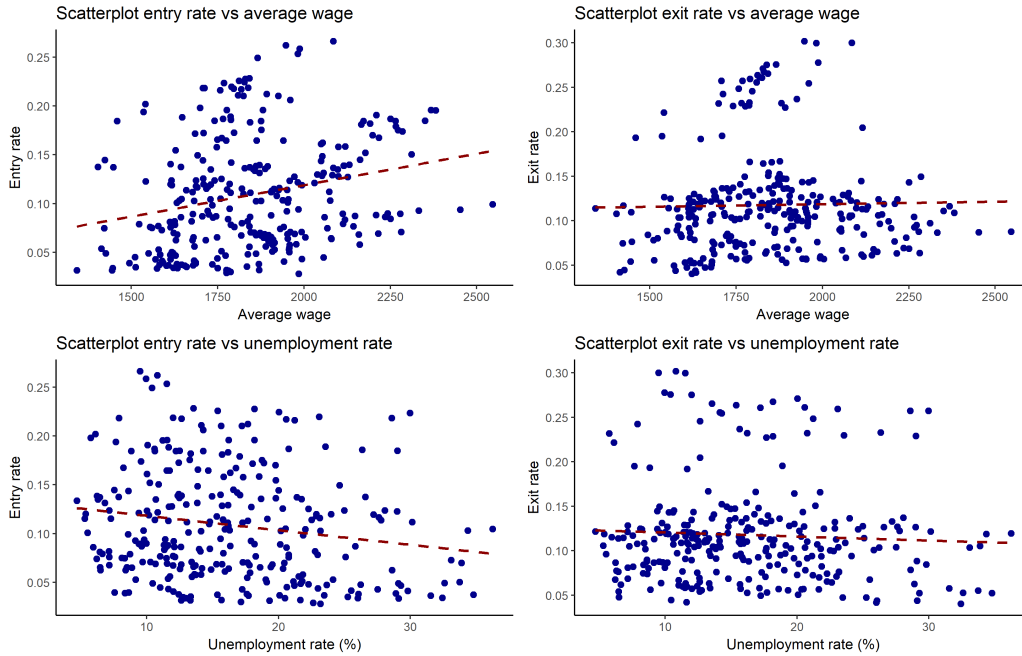
[Figure 3](#) represents how the values for this entry and exit rates are distributed across the different regions of Spain. From these two maps we can obtain 2 main conclusions in order to understand the nature of inter-regional labor mobility in Spain.

The first finding that we obtain is that "poor" regions as Extremadura or Andalusia display higher exit rates than entry rates, while "rich" regions as Madrid or Navarre show the opposite behavior. Thus, we could say that there is a process of workers moving from "poor" regions to "rich" regions (However, as I could not access to the micro-data, this statement cannot be confirmed).

The second outcome that we have from these maps is that there are two regions with extremely high entry and exit rates in comparison with the rest. These regions are Castilla-La-Mancha and La Rioja. Apparently, there is no reason behind these regions having that extremely high rates and not others, as both regions are relatively different in terms of population and specially of geographical area, as La Rioja is the second with the lower surface and Castilla-La-Mancha the third region with higher geographical area ([Instituto Geográfico Nacional, 2005](#)). However, these two regions are the ones producing the higher quantity of wine as a share of its GDP. It is important to note that the wine industry is very stationary, so it generates huge flows of temporary workers that may be resulting in these high values for the entry and exit ratios ([Ministerio de Agricultura, Pesca y Alimentación, 2023](#)).

Finally, before moving to the analysis for the different economic sectors and educational levels, [Figure 4](#) presents four scatter plots regarding the relationship between entry rates and exit rates against the average wage and unemployment rate.

Figure 4: Scatter plots of Entry and Exit rates against average wage and unemployment rate



As we appreciate the relationship between these variables seems stronger when looking at the entry rates instead of exit rates. It may seem counter intuitive that we have the same relationship for both the entry and exit rates (while much more attenuated for the exit rates), however, these are only graphical inspections, we cannot reach any conclusion before taking into account both the temporal and regional dimensions of this dataset.

5.2 By economic sector

In this subsection I will be focusing on the time evolution of the average entry and exit rates of the 4 different economic sectors and its regional distribution. The 4 economic sectors that the Spanish National Statistics Institute (INE) considers are: Agriculture, Construction, Industry and Service sector. The

idea is to see whether one sector displays higher entry and exit rates and, therefore, could be more responsible than others in driving inter-regional labor mobility in Spain.

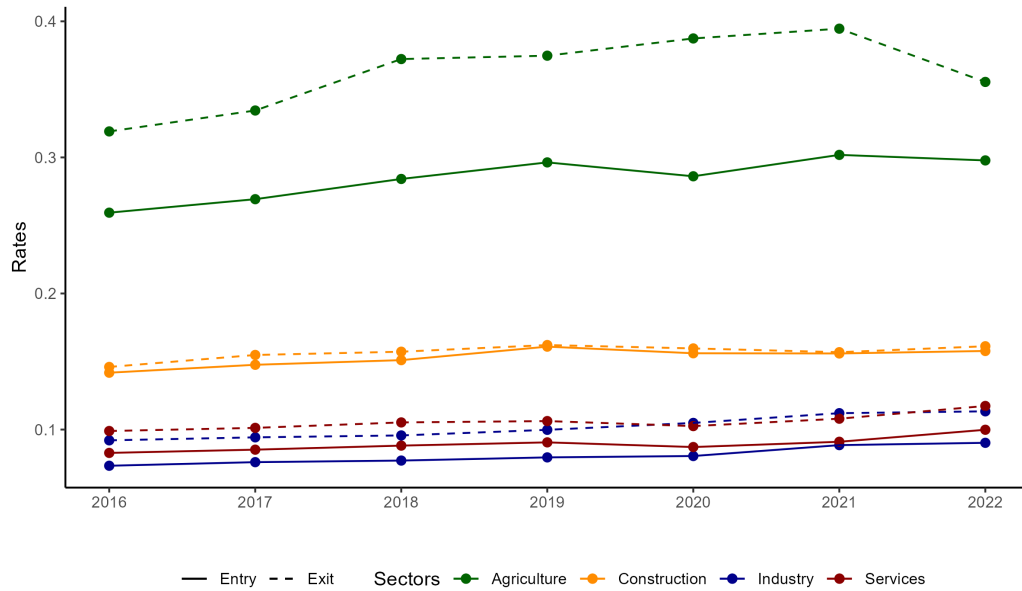
Table 2: Summary statistics of mobility rates by economic sector.

	mean	sd	min	max
Entry rate agriculture	0.285	0.185	0.0362	0.7284
Entry rate construction	0.153	0.0758	0.0207	0.2837
Entry rate industry	0.0808	0.0579	0.0184	0.236
Entry rate services	0.0893	0.0463	0.0271	0.1934
Exit rate agriculture	0.3626	0.2139	0.0661	0.8847
Exit rate construction	0.1568	0.0938	0.0242	0.4198
Exit rate industry	0.1017	0.092	0.0169	0.4499
Exit rate services	0.1056	0.0487	0.0415	0.2661
Observations	119			

Note: Self-elaborated from SEPE data.

Table 2 shows that the economic sector with higher entry and exit rates (on average) is the agricultural. Indeed, its entry ratio almost is 3 times higher than the national average and, in the case of the exit rate, it is more than 3 times higher. The entry and exit rates for the agricultural sector are twice the ones for construction and around 3.5 times higher than the ones for the industrial and service sector.

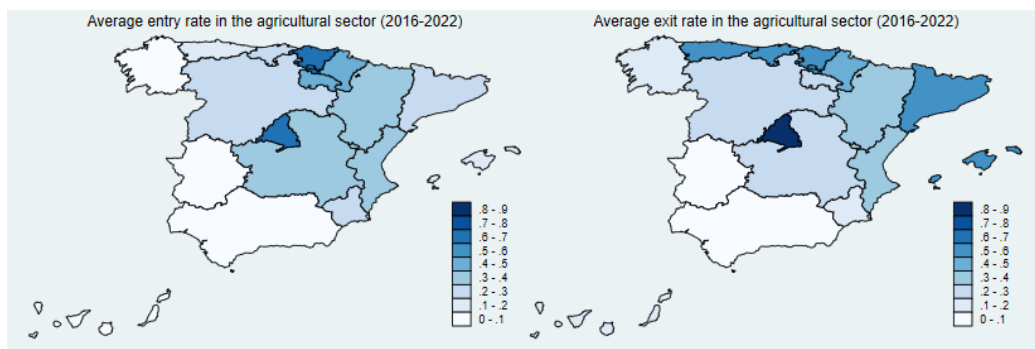
Figure 5: Average entry and exit rates by economic sector and year.



Source: Self-elaborated from SEPE data.

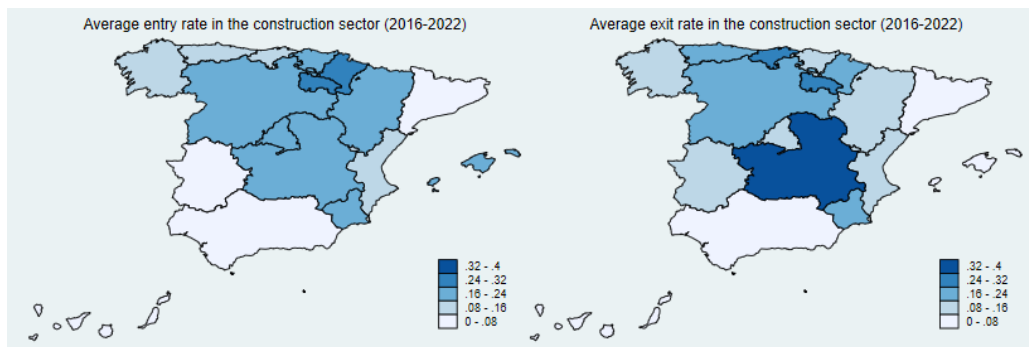
Figure 5 confirms what we have seen previously. As we perceive from the graph, it is the agricultural sector the one having higher entry and exit rates during the whole time span (2016-2022).

Figure 6: Entry and exit rates in the agricultural sector by region (2016-2022).



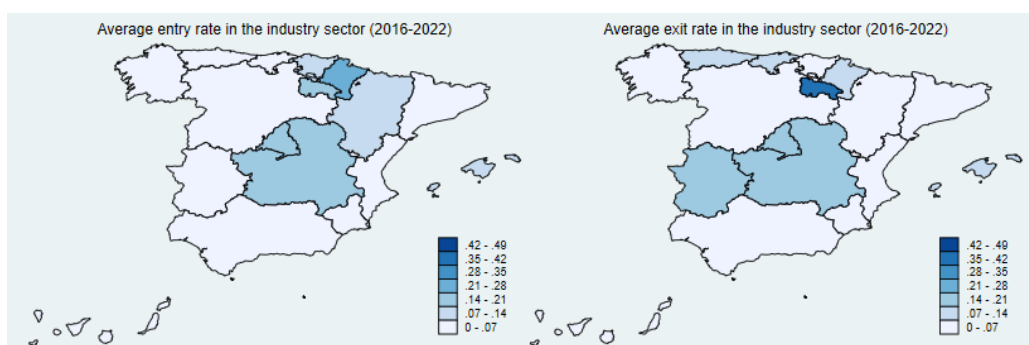
Source: Self-elaborated from SEPE data.

Figure 7: Entry and exit rates in the construction sector by region (2016-2022).



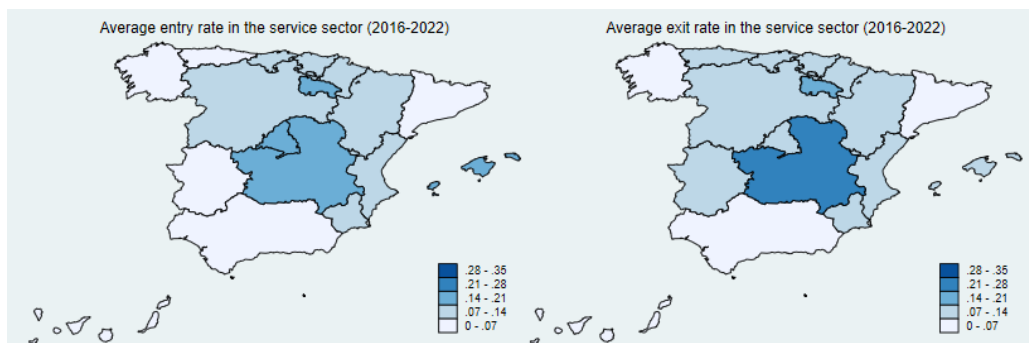
Source: Self-elaborated from SEPE data.

Figure 8: Entry and exit rates in the industrial sector by region (2016-2022).



Source: Self-elaborated from SEPE data.

Figure 9: Entry and exit rates in the service sector by region (2016-2022).



Source: Self-elaborated from SEPE data.

Figures 6, 7, 8 and 9 represent how the entry and exit rates for the different economic sectors are distributed across regions and, as we should expect for what we have seen before, the map yielding higher values for almost all regions (both in entry and exit rates) is the one for agriculture (different scales have been used in order to distinguish more precisely the regional differentials in entry and exit rates).

5.3 By educational level

I will separate the educational level of the Spanish population in 3 different groups; those having secondary education or lower, those having high school (*bachillerato*) or vocational studies (*ciclo formativo*) and those having higher education.

Table 3: Summary statistics of mobility rates by educational level.

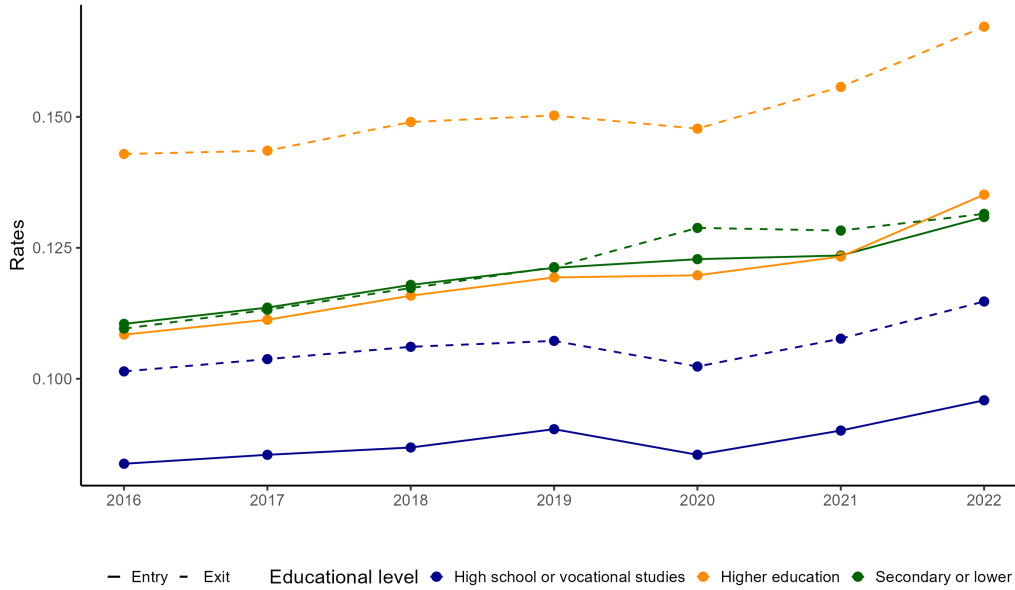
	mean	sd	min	max
Entry rate secondary or lower	0.1201	0.0695	0.0222	0.278
Entry rate high school/vocational studies	0.0883	0.0503	0.0275	0.2021
Entry rate higher education	0.119	0.0602	0.0392	0.3121
Exit rate secondary or lower	0.1214	0.0586	0.0384	0.3061
Exit rate high school/vocational studies	0.1062	0.0598	0.0377	0.2796
Exit rate higher education	0.1509	0.0712	0.0458	0.3437
Observations	119			

Note: Self-elaborated from SEPE data.

Table 3 represents the main statistics for the entry and exit rates for the different educational levels. From its results we are able to discern that the group mobilizing the most is the one of highly educated workers, because not only they have higher exit rates and almost identical entry rates as the low

educated ones, but they also have the higher maximum and minimum values for both entry and exit rates.

Figure 10: Mobility rates by educational level and year.

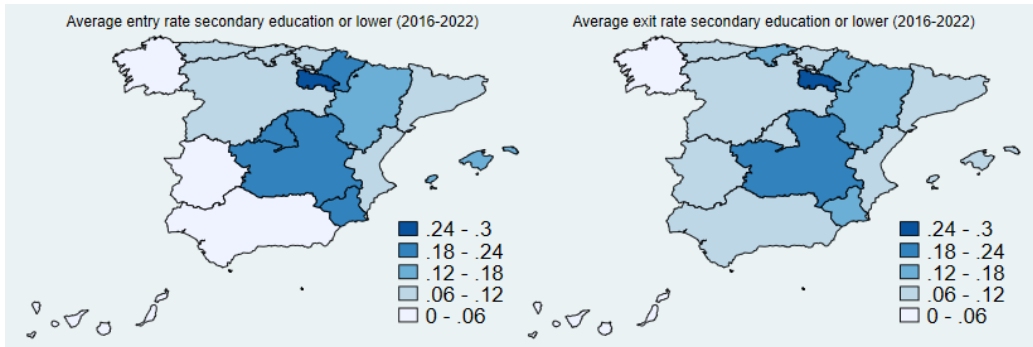


Note: Self-elaborated from SEPE data.

In [Figure 10](#) we can clearly verify the statement I have just said. Although the series for the highly educated workers move similar to the ones of secondary education or lower when looking at the entry rates, the yellow dashed line representing the exit rates of workers with higher education clearly moves above the others.

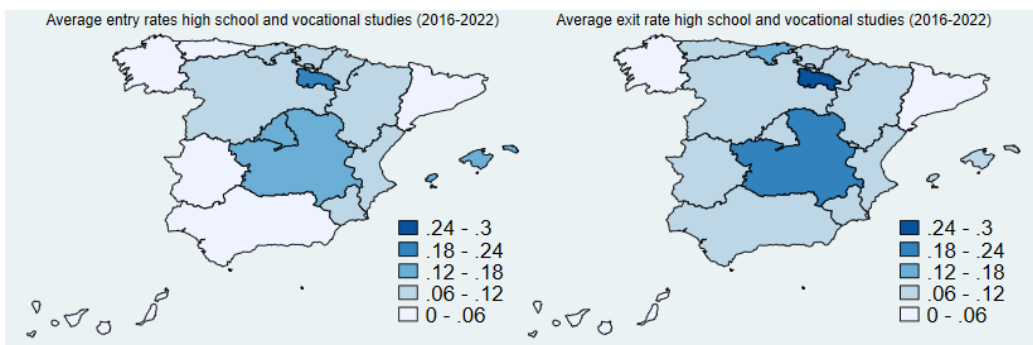
Figures [11](#), [12](#) and [13](#) represents how these entry and exit rates are distributed for the different autonomous communities and educational levels. We clearly appreciate how, specially for the occupied with the higher formation the entry rates (relative to their exit rates) are higher for more economic "developed" regions such as Madrid, Catalonia or Navarre and the opposite happens with the exit rates (regions as Extremadura, Andalusia or Castilla-La-Mancha have higher exit rates than entry rates).

Figure 11: Average entry and exit rates secondary education or lower (2016-2022).



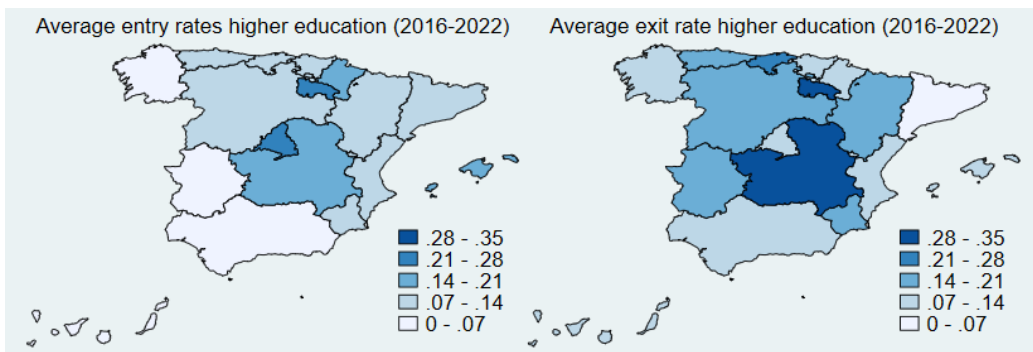
Source: Self-elaborated from SEPE data.

Figure 12: Average entry and exit rates high school and vocational education (2016-2022).



Source: Self-elaborated from SEPE data.

Figure 13: Average entry and exit rates higher education (2016-2022).



Source: Self-elaborated from SEPE data.

6 Methodology

Considering the data characteristics described in [Section 4](#), the main objective of this section is to explain the decisions that I have taken in order to control for both the regional dimension and the temporal dimension of my dataset. Furthermore, I will present the regressions that will be estimated and presented in [Section 7](#).

6.1 Regional control

The procedure that I have taken in order to control for the cross-sectional dimension of my data is the one as follows:

Firstly, I tried to carry out the regressions using regional fixed effects (one dummy variable for each autonomous community), however, it absorbed a significant amount of variability between regions. Therefore, the best way I found to control for the different regions and let the model have more variability was to divide Spain into 6 regional groups:

- **South:** Andalusia and Extremadura.
- **Center:** Both Castilles and Madrid.
- **East:** Valencian Community and Murcia.
- **Northwest:** Galicia, Cantabria and Asturias.
- **Northeast:** Basque Country, La Rioja, Aragon, Navarre and Catalonia.
- **Islands:** Canary Islands and Balearic Islands.

So, what I will include in my estimated models is one dummy variable for each regional group leaving the *Islands'* group as the reference one in order to avoid the dummy trap.

6.2 Temporal control

In order to control for the temporal dimension, I created a fictional variable called t that will take the value of 1 in 2006, 2 in 2007 and so on. The explanation behind carrying out this control and not taking a dummy for each year is that I want to control for the tendency that most of my variables have and, also, because most of the labor migration has a long-term nature. Thus, controlling for the tendency instead of yearly effects captures better the long term perspective of labor migration. As the idea is to control for the tendency, t^2 will be also included.

6.3 Estimated equations

My idea will be to analyse separately the factors affecting the entry rate and exit rate. Based on equations (4) and (5) I will be estimating the following two models (using OLS):

$$Entry_{i,t} = \beta_0 + \beta_1 W_{i,t-1} + \beta_2 u_{i,t-1} + \beta_3 R_{i,t-1} + \beta_4 \Phi_{i,t} + \beta_5 \Gamma_{i,t} + \beta_6 A_i + \beta_7 t + \beta_8 t^2 + e_{i,t} \quad (6)$$

$$Exit_{i,t} = \beta_0 + \beta_1 W_{i,t-1} + \beta_2 u_{i,t-1} + \beta_3 \Phi_{i,t} + \beta_4 \Gamma_{i,t} + \beta_5 A_i + \beta_6 t + \beta_7 t^2 + e_{i,t} \quad (7)$$

Equations (6) and (7) will be the models that I will estimate for both the entry and exit rate respectively, where $W_{i,t-1}$ is the average monthly salary (in thousands of euros) in region i in year $t - 1$, $u_{i,t-1}$ is the unemployment rate (in base 1) of region i in year $t - 1$ and $R_{i,t-1}$ is the average renting price of one square meter in region i in year $t - 1$.

$\Phi_{i,t}$ is a vector composed by 3 variables that tries to capture structural factors of each region (factors non related with the economic sectors of the region).

$$\Phi \in \left\{ \begin{array}{c} Temporary_{i,t} \\ Gini_{i,t} \\ OccHigher_{i,t} \end{array} \right\}$$

Where $Temporary_{i,t}$ represents the percentage of temporary workers of region i in t , $Gini_{i,t}$ is the Gini Index of region i in t ([Appendix C](#) explains how the Gini Index has been computed) and $OccHigher_{i,t}$ is the percentage of workers with higher education of region i in t .

$\Gamma_{i,t}$ is a vector composed by 2 variables that tries to capture economic factors of each region (factors related with the economic sectors of the region).

$$\Gamma \in \left\{ \begin{array}{c} OccAgriculture_{i,t} \\ NonTourismServ_{i,t} \end{array} \right\}$$

Where $OccAgriculture_{i,t}$ represents the share workers in the agricultural sector of region i in t , $NonTourism_{i,t}$ tries to approximate the weight of all the services excluding tourism in the GDP of region i in t ([Appendix C](#) explains how this variable has been computed).

$A_{i,t}$ is a vector composed by 5 regional groups as mentioned in [Section 6.1](#).

$$A \in \left\{ \begin{array}{c} South_i \\ Center_i \\ East_i \\ NorthWest_i \\ NorthEast_i \end{array} \right\}$$

Finally, the unique difference between equations (6) and (7) is that the average renting price for the square meter appears in the *Entry* equation but not in the *Exit* one. In this part, I will provide the economic intuition behind this decision, nevertheless, [Appendix D](#) shows the model for the exit rate including the *rent* variable.

The explanation of assuming that housing prices may have an effect on the entry rates of workers in a region but not on workers deciding whether to migrate from that region is the following: Imagine a worker in region i that is paying a rent there and that has the same job offer in region j and k , where the only difference is that renting prices are higher in region j compared to region k . When computing its discounted returns of migrating ([Equation \(1\)](#)) it is clear that, if the worker has decided to migrate, he or she will move to region k . Thus, in the case of deciding where to move, lower renting costs will result in higher entry rates *ceteris paribus*. However, if the worker has not decided whether to migrate or not, it is not that clear that this decision is mainly driven by the differential of monetary costs (housing costs), as people also take into account other non-monetary costs of leaving a region (as mentioned in [Section 3](#)) that reflect some kind of opportunity costs.

In other words, if a worker has previously decided to move, differentials in the renting costs of the regions will affect their migrating decisions *ceteris paribus*. However, if the decision to exit the region has not been taken, renting prices from the region of origin are not crucial, as there are probably other (in most cases non-monetary) costs such as leaving family or friends behind ([Maza and Villaverde, 2004](#)).

7 Empirical Findings

The goal of this part is to present the outcomes from estimating equations (6) and (7) and discuss the results that have been obtained. Nonetheless, I will not only present the equations where both the structural controls ($\Phi_{i,t}$) and sectorial controls ($\Gamma_{i,t}$) are included, but the evolution of the results when accumulating each control variable.

To clarify, I will present models (6) and (7) each one with 6 columns; the first one without any control, the 3 next, accumulating the "structural" controls and the last 2 adding the economic controls.

Firstly, I will present the results for the overall entry and exit rates of the region and, afterwards, I will present the same results for the economic sector and the educational level having the higher mobility rates (the agricultural sector and the workers with higher education respectively).

Table 4: OLS estimates for the overall entry rates

	(1)	(2)	(3)	(4)	(5)	(6)
	Entry rate	Entry rate	Entry rate	Entry rate	Entry rate	Entry rate
Wage (1 Lag)	0.060** (0.029)	0.030 (0.031)	0.031 (0.031)	0.173*** (0.043)	0.162*** (0.044)	0.153*** (0.044)
Unemployment (1 Lag)	-0.192** (0.077)	-0.236*** (0.078)	-0.225*** (0.078)	-0.255*** (0.076)	-0.250*** (0.076)	-0.220*** (0.076)
Rent (1 Lag)	-0.007*** (0.002)	-0.008*** (0.002)	-0.007*** (0.002)	-0.007*** (0.002)	-0.009*** (0.002)	-0.006*** (0.002)
Temporary		-0.003*** (0.001)	-0.003*** (0.001)	-0.002** (0.001)	-0.002** (0.001)	-0.003*** (0.001)
Gini			-0.007 (0.005)	-0.004 (0.005)	-0.004 (0.005)	-0.001 (0.005)
OccHigher				-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)
OccAgriculture					-0.279 (0.240)	-0.363 (0.241)
NonTourismServ						-0.305** (0.133)
Center	0.043*** (0.012)	0.032** (0.012)	0.031** (0.012)	0.032*** (0.012)	0.042*** (0.014)	0.045*** (0.014)
East	0.043*** (0.013)	0.053*** (0.013)	0.054*** (0.013)	0.051*** (0.013)	0.074*** (0.023)	0.081*** (0.023)
NorthEast	0.016 (0.012)	0.005 (0.013)	0.003 (0.013)	0.011 (0.013)	0.021 (0.015)	0.003 (0.017)
NorthWest	-0.054*** (0.012)	-0.063*** (0.012)	-0.065*** (0.012)	-0.047*** (0.012)	-0.038*** (0.014)	-0.042*** (0.014)
South	-0.031*** (0.011)	-0.012 (0.013)	-0.013 (0.013)	-0.007 (0.012)	0.007 (0.017)	0.029 (0.020)
t	0.004 (0.005)	0.006 (0.005)	0.012* (0.007)	0.014** (0.007)	0.014** (0.007)	0.014** (0.007)
t^2	-0.000 (0.000)	-0.000 (0.000)	-0.001* (0.000)	-0.001** (0.000)	-0.001** (0.000)	-0.001* (0.000)
Constant	0.068* (0.038)	0.206*** (0.064)	0.390*** (0.148)	0.181 (0.150)	0.220 (0.154)	0.248 (0.153)
N	245	245	245	245	245	245
R^2	0.486	0.501	0.505	0.545	0.547	0.558

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Self-elaborated from SEPE, INE and Idealista data

Table 5: OLS estimates for the overall exit rates

	(1)	(2)	(3)	(4)	(5)	(6)
	Exit rate	Exit rate	Exit rate	Exit rate	Exit rate	Exit rate
Wage (1 Lag)	-0.170*** (0.024)	-0.208*** (0.027)	-0.186*** (0.026)	-0.096** (0.041)	-0.049 (0.044)	-0.048 (0.042)
Unemp (1 Lag)	0.072 (0.082)	0.041 (0.082)	0.055 (0.079)	0.049 (0.078)	0.003 (0.078)	0.029 (0.075)
Temporary		-0.003*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)	-0.004*** (0.001)	-0.005*** (0.001)
Gini			-0.024*** (0.005)	-0.022*** (0.005)	-0.020*** (0.005)	-0.012** (0.005)
OccHigher				-0.003*** (0.001)	-0.002** (0.001)	-0.002* (0.001)
OccAgriculture					0.584*** (0.204)	0.135 (0.218)
NonTourismServ						-0.571*** (0.120)
Center	0.104*** (0.011)	0.094*** (0.011)	0.086*** (0.011)	0.089*** (0.011)	0.060*** (0.015)	0.066*** (0.014)
East	0.043*** (0.014)	0.056*** (0.014)	0.054*** (0.014)	0.052*** (0.013)	0.000 (0.022)	0.032 (0.022)
NorthEast	0.101*** (0.012)	0.093*** (0.012)	0.079*** (0.012)	0.088*** (0.012)	0.057*** (0.016)	0.023 (0.017)
NorthWest	0.054*** (0.011)	0.047*** (0.011)	0.033*** (0.011)	0.047*** (0.012)	0.019 (0.015)	0.010 (0.015)
South	-0.011 (0.011)	0.010 (0.013)	0.006 (0.013)	0.008 (0.013)	-0.022 (0.016)	0.033* (0.020)
t	0.004 (0.005)	0.004 (0.005)	0.026*** (0.007)	0.026*** (0.006)	0.023*** (0.006)	0.018*** (0.006)
t^2	0.000 (0.000)	0.000 (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001** (0.000)
Constant	0.318*** (0.037)	0.476*** (0.066)	1.117*** (0.149)	0.994*** (0.153)	0.870*** (0.157)	0.913*** (0.151)
N	272	272	272	272	272	272
R^2	0.336	0.356	0.408	0.425	0.443	0.488

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Self-elaborated from SEPE and INE data

Table 4 and Table 5 show the results for the estimations of the overall entry rate and exit rate models respectively. I will comment separately each table and, then, I will compare both findings.

The main conclusion that we can obtain from Table 4 is how the 3 main variables that may be resulting in a *pull effect* (average wage, unemployment rate and renting costs) have the sign that we expected them to have in equation (4) and that they are quite significant even after adding all controls (showing a sign of robustness to addition of regional controls). Specially, the unemployment rate and the renting costs are significant (and negative) at 1% in all six columns, while the average monthly wage (in thousand of euros) becomes significant at 1% when all the structural controls have been included (column 4) and remains still significant after adding the two sectorial controls. Nonetheless, even though not being significant in all columns, it is always positive, giving support of the positive effect that we expected it to have.

Moving on to Table 5 we are still finding the expected results for the 2 main variables as expressed in equation (5) meaning that, even though not always being significant, higher wages are expected to diminish the expected exit rate of a region while, higher unemployment rates, tend to increase the exit rates of a certain region.

In addition, what we can appreciate by comparing these 2 tables is that, according to their R^2 , the model works relatively better for the entry rates than for the exit rates (Nevertheless, both the R^2 of both models are quite high and yielding similar values).

Probably, the most counter-intuitive part of these results are the interpretation of the 5 control variables. However, the aim for these variables is not that they have a logical sign as I do not expect them to be driven causes of labor mobility or to generate a *pull effect*. The goal

is to control for region-specific economic factors so that my results show robustness in expected sign for the main variables during the accumulation process of controls. Indeed, in most cases (and as I will show afterwards) the majority of controls have the same sign both in the entry and exit equation, meaning that they are capturing characteristics common for regions with both high (or low) entry and exit rates. This is because, as I showed in [Figure 1](#), there is a very strong correlation between these variables.

7.1 By economic sector

Before presenting the results from estimating the two models for the industrial sector, I would like to note that I will not be using the overall average monthly wage of the region, but the annual net income of the specific economic sector or educational level, in order to provide more precise conclusions.

The way I got these annual net income data is from the micro-data of the Life Conditions Survey (*Encuesta de Condiciones de Vida*) from the *INE* and it is explained in detail in [Appendix B](#). These net annual incomes will be expressed in hundreds of thousands of euros in order to have an easier coefficient to interpret.

Table 6: OLS estimates for the agricultural sector entry rates

	(1)	(2)	(3)	(4)	(5)	(6)
	Entry rate	Entry rate	Entry rate	Entry rate	Entry rate	Entry rate
Net Income (1 Lag)	0.014 (0.176)	0.079 (0.188)	0.060 (0.190)	-0.200 (0.164)	-0.098 (0.153)	-0.093 (0.153)
Unemployment (1 Lag)	-1.507*** (0.452)	-1.230** (0.531)	-1.247** (0.533)	-0.855* (0.451)	-0.608 (0.420)	-0.747* (0.449)
Rent (1 Lag)	0.013** (0.005)	0.009 (0.006)	0.008 (0.006)	-0.010 (0.006)	-0.034*** (0.008)	-0.038*** (0.009)
Temporary		-0.007 (0.007)	-0.007 (0.007)	-0.006 (0.006)	-0.006 (0.006)	-0.003 (0.006)
Gini			0.015 (0.019)	0.008 (0.016)	0.016 (0.015)	0.012 (0.016)
OccHigher				0.016*** (0.002)	0.010*** (0.003)	0.009*** (0.003)
OccAgriculture					-4.276*** (0.996)	-4.336*** (1.000)
NonTourismServ						0.421 (0.473)
Center	0.209*** (0.038)	0.182*** (0.047)	0.183*** (0.047)	0.111*** (0.041)	0.228*** (0.047)	0.233*** (0.047)
East	0.102* (0.053)	0.121** (0.056)	0.121** (0.056)	0.099** (0.047)	0.450*** (0.093)	0.451*** (0.093)
NorthEast	0.140*** (0.043)	0.123*** (0.046)	0.128*** (0.047)	0.004 (0.044)	0.143*** (0.052)	0.172*** (0.061)
NorthWest	-0.071 (0.044)	-0.088* (0.047)	-0.083* (0.048)	-0.247*** (0.047)	-0.130** (0.051)	-0.123** (0.052)
South	0.005 (0.048)	0.027 (0.053)	0.027 (0.053)	-0.027 (0.045)	0.212*** (0.070)	0.192** (0.073)
t	-0.146 (0.097)	-0.060 (0.131)	-0.085 (0.135)	0.001 (0.114)	0.081 (0.107)	0.046 (0.114)
t^2	0.005 (0.003)	0.002 (0.005)	0.003 (0.005)	-0.001 (0.004)	-0.003 (0.004)	-0.002 (0.004)
Constant	1.455* (0.739)	1.044 (0.847)	0.703 (0.954)	-0.159 (0.811)	-0.592 (0.756)	-0.380 (0.793)
Observations	115	115	115	115	115	115
R^2	0.665	0.668	0.670	0.770	0.806	0.807

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Self-elaborated from SEPE, INE and Idealista data

Table 7: OLS estimates for the agricultural sector exit rates

	(1)	(2)	(3)	(4)	(5)	(6)
	Exit rate	Exit rate	Exit rate	Exit rate	Exit rate	Exit rate
Net Income (1 Lag)	0.096 (0.251)	0.372* (0.211)	0.285 (0.213)	-0.211 (0.184)	-0.226 (0.154)	-0.253* (0.151)
Unemployment (1 Lag)	-3.114*** (0.646)	-0.768 (0.629)	-0.861 (0.622)	-0.703 (0.502)	-0.863** (0.422)	-1.348*** (0.451)
Temporary		-0.049*** (0.007)	-0.046*** (0.007)	-0.032*** (0.006)	-0.017*** (0.006)	-0.009 (0.006)
Gini			0.044* (0.022)	0.023 (0.018)	0.022 (0.015)	0.006 (0.016)
OccHigher				0.019*** (0.002)	0.007** (0.003)	0.005* (0.003)
OccAgriculture					-4.697*** (0.710)	-3.988*** (0.742)
NonTourismServ						1.097** (0.423)
Center	-0.009 (0.055)	-0.161*** (0.050)	-0.146*** (0.050)	-0.161*** (0.040)	0.063 (0.048)	0.068 (0.047)
East	-0.180** (0.073)	0.025 (0.067)	0.024 (0.066)	0.002 (0.054)	0.397*** (0.075)	0.331*** (0.077)
NorthEast	-0.133** (0.062)	-0.215*** (0.053)	-0.194*** (0.053)	-0.291*** (0.045)	-0.051 (0.052)	0.010 (0.056)
NorthWest	-0.114* (0.061)	-0.158*** (0.051)	-0.136*** (0.051)	-0.260*** (0.045)	-0.012 (0.053)	0.002 (0.052)
South	-0.085 (0.069)	0.095 (0.063)	0.090 (0.062)	0.000 (0.051)	0.251*** (0.057)	0.146** (0.069)
t	-0.168 (0.142)	0.413*** (0.143)	0.316** (0.150)	0.265** (0.121)	0.167 (0.103)	0.027 (0.113)
t^2	0.005 (0.005)	-0.016*** (0.005)	-0.012** (0.005)	-0.011** (0.004)	-0.007* (0.004)	-0.002 (0.004)
Constant	2.258** (1.075)	-0.742 (0.985)	-1.635 (1.074)	-1.650* (0.865)	-0.823 (0.737)	0.033 (0.789)
Observations	115	115	115	115	115	115
R^2	0.455	0.632	0.645	0.772	0.841	0.851

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Self-elaborated from SEPE and INE data

Table 6 and Table 7 display the results obtained from estimating the models presented in Section 6.3, but using the entry and exit rates for the agricultural sector (as expressed in Section 3.1) and the annual net income for the agricultural sector (in hundreds of thousand of euros) instead of the average monthly wage.

The first conclusion that we obtain is that the annual net income in agriculture is not a driving force for labor mobility in the sector. Its coefficient is sometimes positive and sometimes negative both in the entry and the exit and, almost never, statistically significant. The explanation for this result is that, due to the huge temporality and precariousness in this sector, workers migrate because they need to work (most probably they have moved from being unemployed to being employed). Thus, the main driving force should be the probability of finding a job and not that much, the earnings they will obtain (as I have just said many of them would most probably move from not working (and obtaining no income) to start working. However, as I could not access the micro-data, I cannot confirm this).

Following with the the interpretation of the coefficients for the (overall) unemployment rate I got that it is very negative and significant for the entries and negative, but not always significant, for the exits. These outcomes are in line with the explanation I made for the net income results. As I have just said, most probably, the main driving force for this type of workers is the fact that they move from being unemployed to being employed. Therefore, the higher the probability of finding a job in a region is (the lower unemployment is) the higher the expected entry rate of this type of workers is.

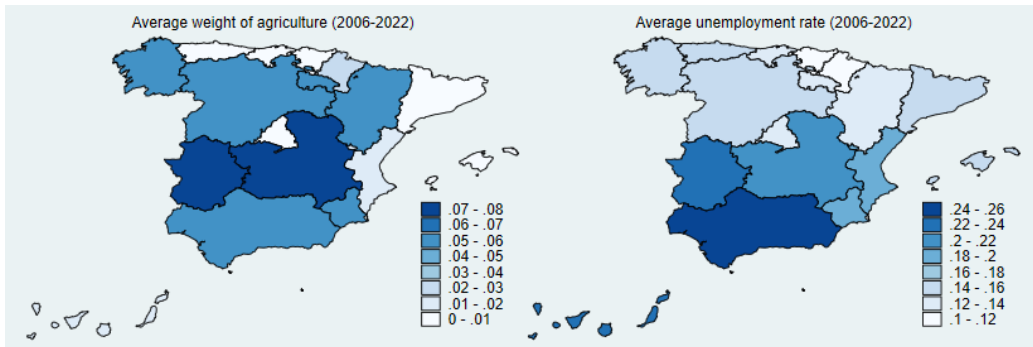
In order to explain why do we obtain the same sign both in the entry and exit rates estimations it is important to remember the equations used to obtain

the entry and exit rates of an economic sector presented in [Section 3.1](#):

$$\text{Entry rate}_{i,j,t} = \frac{\text{Entries}_{i,j,t}}{\text{Stays}_{i,j,t} + \text{Entries}_{i,j,t}} \quad \text{Exit rate}_{i,j,t} = \frac{\text{Exits}_{i,j,t}}{\text{Stays}_{i,j,t} + \text{Exits}_{i,j,t}}$$

Therefore (and as I will show now), regions with high unemployment are supposed to be more "developed" regions in terms of labor market and therefore we will expect them to have a lower weight of the agricultural sector in their GDP and therefore a lower amount of workers staying in the region to work in the agriculture (relative to the ones that migrate). For that reason, the denominator of these regions (the ones with lower unemployment) will be lower, resulting in both higher entry and exit rates.

Figure 14: Average agriculture weight and unemployment rate by region (2006-2022)



Source: Self-elaborated from INE data.

Figure 15: Correlation matrix for industrial labor mobility



Source: Self-elaborated from SEPE and INE data.

Figure 14 and Figure 15 give context on which are the regions with more and less agricultural weight, its average unemployment rates and the correlation coefficients between entry and exit rates in the agricultural sector of each region and the relative weight of its agriculture and their unemployment rate. As we appreciate, there is a negative relationship between both entry and exit rates in the agricultural sector and the weight of the agriculture in the region, but this negative relationship is much stronger for the exit rates, supporting my previous statement.

Nevertheless, I would like to stress the fact that, as I could not get the unemployment rate for this group of workers, this variable should be interpreted as a control variable in a way. So, as I have mentioned in the previous section, this could also be an alternative explanation of why do we encounter the same sign both in entry and exit estimations.

In addition to all what I have just presented, two more conclusions can be drawn. The first one is how the renting prices are not a driving force for determining the entry rate, as they half of the time positive and the other half negative and not always statistically significant. This is in line with my previous statements regarding the temporality of the sector. As most of the workers will not stay for long periods of time, living costs are not expected to be a determinant of the labor mobility in agriculture.

The second inference we can make is that, contrary to what we saw for the overall entry and exit models, this specification works very similarly for both the entry rate estimation than for the exit rate estimation according to the R^2 fitness criteria. Furthermore, the values for the R^2 of these estimations are much more larger now than for the general entry and exit rates.

7.2 By educational level

The last results that I will present are the ones for the educational level which generates the higher labor mobility weight: the workers that have obtained higher education.

I will present both the estimations for its entry and exit rates and analyze them briefly as I did with both the overall and the agricultural sector entry and exit rates. In this case, I will be using the annual net income (in hundreds of thousand of euros) computed from the Life Conditions Survey for this specific group as well.

Table 8: OLS estimates for the entry rates of workers with higher education

	(1)	(2)	(3)	(4)	(5)	(6)
	Entry rate	Entry rate	Entry rate	Entry rate	Entry rate	Entry rate
Net Income (1 Lag)	0.555 (0.420)	0.950** (0.428)	0.974** (0.434)	1.162*** (0.431)	0.855** (0.403)	0.869** (0.405)
Unemployment (1 Lag)	-0.675*** (0.191)	-0.332 (0.219)	-0.328 (0.220)	-0.391* (0.217)	-0.369* (0.200)	-0.408* (0.213)
Rent (1 Lag)	-0.006** (0.003)	-0.012*** (0.003)	-0.012*** (0.003)	-0.009** (0.004)	-0.019*** (0.004)	-0.020*** (0.004)
Temporary		-0.008*** (0.003)	-0.008*** (0.003)	-0.008*** (0.003)	-0.007** (0.003)	-0.006** (0.003)
Gini			-0.003 (0.008)	-0.002 (0.008)	0.003 (0.007)	0.002 (0.007)
OccHigher				-0.003** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)
OccAgriculture					-2.012*** (0.453)	-2.026*** (0.455)
NonTourismServ						0.117 (0.216)
Center	0.009 (0.018)	-0.030 (0.022)	-0.031 (0.022)	-0.021 (0.022)	0.045* (0.025)	0.045* (0.025)
East	-0.039* (0.022)	-0.022 (0.022)	-0.023 (0.022)	-0.022 (0.022)	0.142*** (0.042)	0.142*** (0.042)
NorthEast	-0.029 (0.018)	-0.051*** (0.019)	-0.052*** (0.019)	-0.031 (0.021)	0.039 (0.025)	0.046 (0.029)
NorthWest	-0.077*** (0.018)	-0.099*** (0.019)	-0.100*** (0.019)	-0.073*** (0.022)	-0.020 (0.023)	-0.018 (0.024)
South	-0.029 (0.021)	-0.011 (0.021)	-0.012 (0.021)	-0.007 (0.021)	0.105*** (0.032)	0.099*** (0.034)
t	-0.046 (0.046)	0.072 (0.060)	0.078 (0.062)	0.071 (0.061)	0.072 (0.056)	0.062 (0.059)
t^2	0.001 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)
Constant	0.533 (0.377)	-0.115 (0.427)	-0.054 (0.455)	0.014 (0.445)	0.036 (0.410)	0.092 (0.425)
Observations	119	119	119	119	119	119
R^2	0.444	0.485	0.485	0.513	0.591	0.592

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Self-elaborated from SEPE, INE and Idealista data

Table 9: OLS estimates for the exit rates of workers with higher education

	(1)	(2)	(3)	(4)	(5)	(6)
	Exit rate	Exit rate	Exit rate	Exit rate	Exit rate	Exit rate
Net Income (1 Lag)	-2.122*** (0.365)	-1.960*** (0.367)	-1.790*** (0.380)	-0.831** (0.387)	-0.964** (0.465)	-0.933* (0.472)
Unemployment (1 Lag)	-0.254 (0.242)	-0.494* (0.265)	-0.459* (0.264)	-0.492** (0.237)	-0.516** (0.242)	-0.463* (0.270)
Temporary		0.006** (0.003)	0.005* (0.003)	0.003 (0.003)	0.004 (0.003)	0.003 (0.004)
Gini			-0.016 (0.010)	-0.012 (0.009)	-0.012 (0.009)	-0.010 (0.009)
OccHigher				-0.006*** (0.001)	-0.007*** (0.001)	-0.007*** (0.002)
OccAgriculture					-0.250 (0.479)	-0.300 (0.493)
NonTourismServ						-0.112 (0.248)
Cneter	0.123*** (0.020)	0.139*** (0.021)	0.132*** (0.022)	0.134*** (0.019)	0.147*** (0.032)	0.146*** (0.032)
East	0.017 (0.026)	-0.007 (0.028)	-0.009 (0.028)	-0.012 (0.025)	0.009 (0.048)	0.014 (0.049)
NorthEast	0.051** (0.022)	0.060*** (0.022)	0.052** (0.022)	0.091*** (0.021)	0.103*** (0.032)	0.096*** (0.035)
NorthWest	0.032 (0.021)	0.040* (0.021)	0.032 (0.021)	0.084*** (0.021)	0.094*** (0.030)	0.092*** (0.030)
South	0.038 (0.025)	0.014 (0.027)	0.012 (0.027)	0.021 (0.024)	0.036 (0.037)	0.045 (0.042)
t	-0.117** (0.054)	-0.179*** (0.061)	-0.143** (0.065)	-0.103* (0.059)	-0.115* (0.063)	-0.099 (0.073)
t^2	0.005** (0.002)	0.007*** (0.002)	0.005** (0.002)	0.004* (0.002)	0.004* (0.002)	0.004 (0.003)
Constant	1.272*** (0.439)	1.533*** (0.450)	1.803*** (0.478)	1.557*** (0.431)	1.646*** (0.465)	1.544*** (0.518)
Observations	119	119	119	119	119	119
R^2	0.355	0.379	0.394	0.518	0.519	0.520

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Self-elaborated from SEPE and INE data

Firstly, when looking at the coefficients obtained for the net annual rent (in hundreds of thousands of euros), we can appreciate very interesting results. Clearly, for this type of workers, their annual rent becomes a very strong driving force both in the exit and entry rate. For the entry rate, it is always positive and, almost always, significant, meaning that the higher the remuneration these workers expect to receive in a region, the greater the expected entry rates for the highly educated workers in that region are. In contrast, the lower the salary they are getting in a region, the larger the probability of this workers to migrate from that region to another. Therefore, we can conclude that, opposite to what we saw before with the agricultural sector, now the expected salary is clearly a reason for workers to migrate.

This makes sense, since the type of workers found in this group are very distinct from those working in agriculture as the kind of jobs these workers are at, might probably be the ones of greater value added in the economy. So, they do not care that much for securing work but for obtaining bigger earnings (I do not have data regarding unemployment rates for different educational levels, but these workers most probable will have the lowest ones). Focusing on the interpretation of the coefficients for the unemployment rate the following inferences can be done:

As we appreciate in both models, these coefficients are negative and sometimes significant, meaning that higher unemployment rates lead to both lower expected entry and exit rates. In order to understand the logic behind this, it is very important to remember that, as this variable does not represent the unemployment rate for this group of workers, it captures regional characteristics. Thus, as I said in the previous section, regions with higher unemployment rates tend to be associated with less "developed" regions in economic terms. Therefore, we could say, in light of these results,

that more economic "developed" regions not only "import" highly educated workers but they also "export" them.

The rationale behind this is that, as they have more jobs requiring highly educated workers, they offer more competitive salaries to workers from other regions resulting in these regions having higher entry rates. Furthermore, these regions also have a more advance educational system for superior education, so many of the high educated workers exiting these regions most probably might be students that recently graduated and that are starting to work in another region. Once again, I cannot confirm this hypothesis by looking at the micro-data, nevertheless, it goes in line with the coefficients obtained in [Table 8](#) and [Table 9](#).

Next, moving on to the renting prices, we find that its coefficients are negative and significant for the entry rate, which is what we would clearly expected as, with great certainty, these labor migrations are of long-term nature.

Finally, comparing the goodness of fit of both models, we encounter something similar to what we found in the agricultural sector: the values obtained for the R^2 are quite similar for both the entry and exit rates (specially when almost all controls have been included). In addition, these values are very similar to the ones obtained for the overall entry and exit rates and considerably lower than the ones for the agriculture.

8 Conclusions

To sum up, I have gathered numerous takeaways during the analysis of the recent evolution of labor mobility in Spain and the econometric analysis performed. Most of them are summarized in this last section:

Firstly, it is very important to mention the role that labor migrations play in a country such as Spain, where regional labor and educational inequalities have been (and still are) present during all its history. We have also seen how this labor mobility was intensified during the great recession and that it is particularly high for the agricultural sector and for the workers that are highly educated.

Using data coming from various sources (*SEPE*, *EPA* or *ECV* among others) I have specified separately two models: one for the entry ratio and another for the exit ratio. I estimated these two models for the overall regional entry and exit rates and also for the agricultural sector and the workers that are highly educated, as these are the two groups of workers resulting in higher mobility rates if we dis-aggregate by economic sector or educational level.

The results that I obtained were that both the average wage and the unemployment rate are crucial factors for driving labor mobility and, indeed, they have opposite effects for both the entry and exit rates, and we should expect. Regarding the entry rate, the renting costs clearly have a negative a significant effect. When looking at the mobility rates for the agriculture and the highly educated workers we reach the following outcomes:

For the agricultural sector, due to its huge temporality and precariousness, what truly matters is the will of securing a job and not really the expected remuneration. Furthermore, due to the high temporality, the renting costs

have an unclear effect for determining the entry rates. When examining the estimations for the entry and exit rates of the highly educated workers we discover that their driving forces are the opposite ones (and it makes sense as the workers constituting this group are completely different from the ones being in the agricultural sector) These workers care more about the expected salary that they will receive rather than the probability of being employed. Moreover, as most of this labor migration is thought to be of long term nature, the renting costs do play now an important role for explaining the expected entry rates.

As a final remark, I would like to mention the unique aspects of my work and which are its limitations. I have analyzed separately the entry and the exits in relative terms (using the entry and exit rates) while most of the empirical works previously done only focus on labor mobility as a whole. In addition, I tried to delve into the driving forces of this entries and exits dis-aggregating by economic sector and educational level, something that I have not seen in previous researches.

However, due to the lack of micro-data and not having data for a long period of time (specially for the economic sectors and academic levels) the econometric analysis can be significantly enhanced and provide us with more precise outcomes.

References

- Angulo, A. and Mur, J. (2005). Geographical labour mobility in spain-a panel data approach.
- Bentolila, S. (2001). Las migraciones interiores en españa. *Documento de trabajo*, 7.
- Borjas, G. J., Bronars, S. G., and Trejo, S. J. (1992). Self-selection and internal migration in the united states. *Journal of Urban Economics*, 32(2):159–185.
- Bover, O. and Arellano, M. (2002). Learning about migration decisions from the migrants: Using complementary datasets to model intra-regional migrations in spain. *Journal of Population Economics*, 15:357–380.
- Chhetri, P., Khan, A., Stimson, R., and Western, J. (2009). Why bother to ‘downshift’? the characteristics and satisfaction of downshifter in the brisbane-south east queensland region, australia. *Journal of Population Research*, 26:51–72.
- Clemente, J., Larramona, G., and Olmos, L. (2016). Interregional migration and thresholds: evidence from spain. *Spatial Economic Analysis*, 11(3):276–293.
- Hamidov, E. T. O. (2022). Causes, social consequences and economic impact of labor migration. *Scientific progress*, 3(2):1064–1068.
- Hamilton, C. and Mail, E. (2003). Downshifting in australia: A sea-change in the pursuit of happiness.
- Instituto Geográfico Nacional (2005). España en gráficos. https://www.ign.es/espmap/graficos_spain_bach/Espana_Graf_01.htm. Accessed on May 1st, 2024.
- Kan, K. (1999). Expected and unexpected residential mobility. *Journal of urban economics*, 45(1):72–96.
- Khwaja, Y. (2002). Should i stay or should i go? migration under uncertainty: a real options approach.
- Liu, L. Q. (2018). *Regional labor mobility in Spain*. International Monetary Fund.
- Llorente Jiménez, I. (2020). Análisis econométrico de la movilidad laboral.

- Maza, A. and Villaverde, J. (2004). Interregional migration in Spain: a semiparametric analysis. *Review of regional studies*, 34(2):156–171.
- Ministerio de Agricultura, Pesca y Alimentación (2023). INFOVI campaña 2023/2024. Informe técnico, Ministerio de Agricultura, Pesca y Alimentación.
- Morrison, P. S. and Clark, W. A. (2011). Internal migration and employment: macro flows and micro motives. *Environment and Planning A*, 43(8):1948–1964.
- Mulhern, A. and Watson, J. (2010). Spanish inter-regional migration: an enigma resolved. *Applied economics letters*, 17(14):1355–1359.
- Pronchev, G., A.P., L., Proncheva, N., and Tretiakova, I. (2019). Social and economic causes of labor migration in contemporary Russia. *Espacios*, 40:13.
- Roca, J. D. L. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.
- Simon, H. A. (1957). Models of man; social and rational.
- Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, 70(5):80–93.

Appendix

A. Labor mobility data from SEPE

As I could not access to the micro data of labor mobility from the SEPE, the only way I had to obtain, at least, its entry and exit rates at a regional level was obtaining them from their Basic Mobility Data bulletins that are published since 2006. These bulletins contained information at regional and province level for entry, exit and mobility rates in addition to data dis-aggregated in sex, educational level or sector of work among other classifications from 2016 onwards.

There are other datasets covering labor mobility in Spain such as the "Labor and geographical mobility statistic" carried out by the Spanish National Statistics Institute or the "Mobility of the labor market coming from tax sources" carried out by the Spanish Tax Agency. The main drawbacks from these databases were that, for the one developed by the **INE**, there is not a clear definition for "labor mobility", they only provides you with information from the time since that person is living in a region. The database for the Spanish Tax Agency is very similar to the one I used, as it gives you the amount of workers entering and exiting from each region but, as it is built by the Spanish Tax Agency, neither Navarre nor the Basque Country appear there as they have their own tax systems.

B. Income data coming from ECV

In this part I will briefly comment how I computed the regional net income data for the 4 economic sectors (agriculture, industry, construction and services) and for the 3 educational levels (secondary or inferior, high school

or vocational studies and higher education).

The first thing I do is (from folder P) to keep only the people receiving a salary (variable PL040A = 3 from 2016 to 2020 / PL040A = 3 from 2021-2023) and then I drop the observations not reporting a salary (PY010N = 0). This is common for both the economic sector and educational level computation. Now I classify the economic sectors and educational level in the following manner:

- Economic sector (Variable PL111A from 2016-2020 and variable PL111AA from 2021-2023)
 - Agriculture: A.
 - Industry: B, C, D and E.
 - Construction: F.
 - Services: G, H, I, J, K, L, M, N, O, P, Q, R, S, T and U.

- Educational level (Variable PE040 from 2016-2020 and variable PE041 from 2021-2023)
 - Secondary education or lower: 0, 100 and 200.
 - High school or vocational studies: 300, 344, 353, 354, 400 and 450.
 - Higher education: 500.

Now, I create the household identifier (variable that I will be using to merge folders P and D and then obtain the regional averages for each economic sector or educational level) defined as the lower rounding of dividing the personal identifier (PB030) between 100. After merging folders P and D using the household identifier, the last step is just to get the averages by economic sector (or educational level) and autonomous community (variable DB040).

C. Computation of independent variables

- GINI INDEX

From the decile wage distribution provided by the *INE* for each region for the period 2006-2022 I perform the following Calculations:

1. I added the average salaries corresponding to each decile.

$$W_{j,t} = \sum_{i=1}^{10} w_{i,j,t} \quad \forall j, t$$

Where j is the regional sub-index, t , the temporal sub-index and i corresponds to the i_{th} decile.

2. I compute the relative weight of each decile average wage with the respect to the total sum of average decile wages:

$$y_{i,j,t} = \frac{w_{i,j,t}}{W_{j,t}} \quad \forall i, j, t$$

and its cumulative frequency:

$$Y_{i,j,t} = \sum_{i=1}^{n \leq i} y_{i,j,t} \quad \forall i, j, t$$

3. Then, i define the share of population that should correspond to each decile in a perfect egalitarian economy:

$$x_{i,j,t} = 0.1 \quad \forall i, j, t$$

and its cumulative frequency:

$$X_{i,j,t} = \sum_{i=1}^{n \leq i} x_{i,j,t} \quad \forall i, j, t$$

4. Finally, the formula for the Gini index for the j th region in year t is:

$$G_{j,t} = \sum_{i=1}^{10} [(X_{i,j,t} + X_{i-1,j,t}) \times (Y_{i,j,t} - Y_{i-1,j,t})] - 1$$

This gives a Gini index that goes from 0 to 1. The one I used in the regressions goes from 0 to 100, so the only thing we need to do is multiply that result by 100.

- **NON TOURISM SERVICES**

The Spanish National Statistics Institute provides information for the regional value added of the following activities according to the classification *CNAE 2009*

(Clasificación Nacional de Actividades Económicas). The data is classified in the following branches of activity:

- A: Agriculture, livestock, forestry, and fishing.
- B-E: Extractive industries, manufacturing industry, supply of electricity, gas, steam and air conditioning, water supply, sanitation activities, waste management and decontamination.
- F: Construction.
- G-I: Wholesale and retail trade, repair of motor vehicles and motorcycles, transport and storage, hospitality.
- J: Information and communications.
- K: Financial and insurance activities.
- L: Real estate activities.

- M-N: Professional, scientific and technical activities, administrative activities and auxiliary services.
- O-Q: Public administration and defense, compulsory social security, education, health activities and social services.
- R-U: Artistic, recreational and entertainment activities, repair of household items and other services.

Thus, as I could not get more precise data, what I considered as *Non Tourism Services* were codes J-U and I added their gross added value and divided it by the total gross added value of the region.

D. Exit model using square meter renting price

In this section I will present the results for the overall exit rate estimation including the variable for the average renting costs. In [Section 6.3](#) I provided the economic intuition of why not to include this variable in the exit model, however, now I will show the results including it and demonstrate how it affects (negatively) the estimated coefficients.

Table 10: OLS estimates for the overall exit rates including renting costs

	(1)	(2)	(3)	(4)	(5)	(6)
	Exit Rate	Exit Rate	Exit Rate	Exit Rate	Exit Rate	Exit Rate
Wage (1 Lag)	-0.033 (0.030)	-0.063** (0.032)	-0.062** (0.031)	0.042 (0.045)	0.036 (0.046)	0.029 (0.046)
Unemployment (1 Lag)	-0.031 (0.080)	-0.075 (0.081)	-0.049 (0.079)	-0.070 (0.078)	-0.068 (0.078)	-0.045 (0.079)
Rent (1 Lag)	-0.013*** (0.002)	-0.013*** (0.002)	-0.012*** (0.002)	-0.012*** (0.002)	-0.013*** (0.002)	-0.011*** (0.002)
Temporary		-0.003*** (0.001)	-0.004*** (0.001)	-0.003** (0.001)	-0.003** (0.001)	-0.003*** (0.001)
Gini			-0.016*** (0.005)	-0.014*** (0.005)	-0.014*** (0.005)	-0.012** (0.005)
OccHigher				-0.003*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)
OccAgriculture					-0.168 (0.248)	-0.232 (0.250)
NonTourismServ						-0.232* (0.139)
Center	0.064*** (0.012)	0.052*** (0.013)	0.052*** (0.012)	0.053*** (0.012)	0.058*** (0.015)	0.061*** (0.015)
East	0.016 (0.013)	0.026* (0.014)	0.028** (0.014)	0.026* (0.013)	0.039 (0.024)	0.045* (0.024)
NorthEast	0.057*** (0.013)	0.046*** (0.013)	0.042*** (0.013)	0.048*** (0.013)	0.054*** (0.016)	0.040** (0.018)
NorthWest	0.013 (0.012)	0.003 (0.012)	-0.001 (0.012)	0.012 (0.013)	0.018 (0.015)	0.014 (0.015)
South	-0.023** (0.011)	-0.004 (0.013)	-0.006 (0.013)	-0.002 (0.013)	0.007 (0.018)	0.024 (0.020)
t	-0.008 (0.005)	-0.006 (0.005)	0.009 (0.007)	0.011 (0.007)	0.011 (0.007)	0.011 (0.007)
t^2	0.001** (0.000)	0.000* (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Constant	0.271*** (0.039)	0.411*** (0.066)	0.854*** (0.150)	0.701*** (0.155)	0.724*** (0.159)	0.746*** (0.159)
Observations	245	245	245	245	245	245
R^2	0.443	0.459	0.483	0.504	0.505	0.511

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Self-elaborated from SEPE, INE and Idealista data

As we see, the coefficients for both unemployment and average wage have completely changed and almost never are significant. Therefore, including the renting costs in the exit rate estimation might result in it capturing other characteristics apart for the living costs as it was supposed to do and that distort the estimated model.