Interregional migration and thresholds: evidence in Spain*

Jesús Clemente Gemma Larramona† Lorena Olmos

University of Zaragoza

January 2013

ABSTRACT

The aim of this paper is to analyze the effects of labor market conditions in origin and destination on interregional migration in Spain over the period 1988-2010. A basic theoretical framework is developed and the implications of the model suggest that the effect of labor market conditions of origin and destination on migration can be different depending on a threshold. In a second step the implications of the model are tested with Spanish data by using a new approach based on the presence of thresholds. We show that interregional migration can be explained by labor market fundamentals if the expected wage gap in origin over destination is below an endogenously determinate value.

JEL Code: R23, C20, J61

Keywords: Interregional migration, Thresholds, Spain.

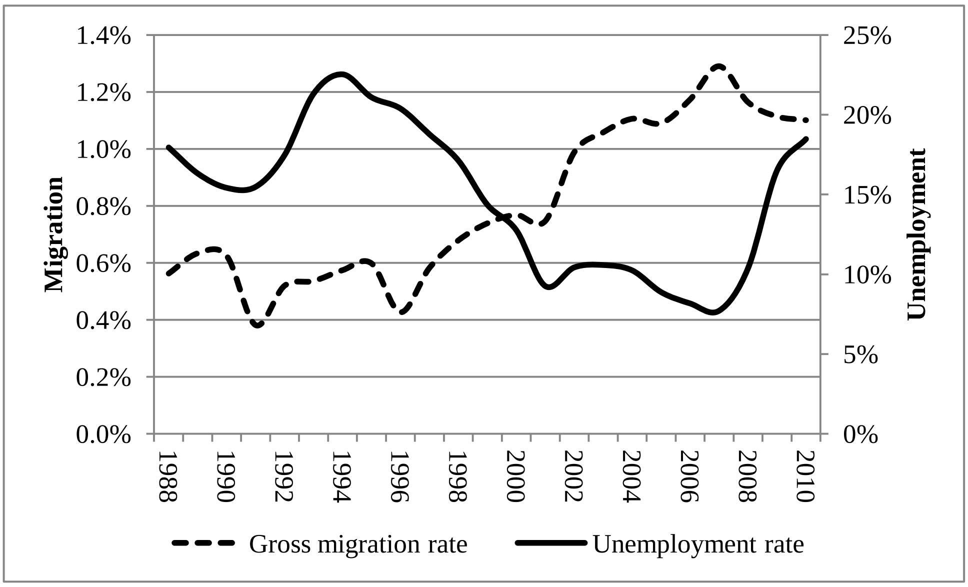
* The authors acknowledge the financial support of the MEC (project X and grant AP2008-02872) and that of the excellence group ADETRE. The usual disclaimer applies.
† Corresponding author. Email: gemmalar@unizar.es

1 Introduction

Empirical literature about migrations has extensively analyzed the determinants of the emigration decision, however, the consensus is far from been reached. From a microeconomic perspective, neoclassical theories explain the migration decision through the difference in salaries between the origin and the destination. With this perspective in mind, most of the papers include salaries as an explanatory variable. Moreover, other differences in the labor market conditions, as in the unemployment rate, are also included in order to capture the probability of finding a job, as Harris and Todaro (1970) first considered. The interregional migrations in Spain are a paradigm in this context. As Bentolila (2001) highlights, Spanish internal migration flows have been low when are compared to other developed European countries and they do not respond clearly to high unemployment in the regions of residence. Bentolila and Dolado (1991) resolve that both real wages and unemployment differences are significant, although they come to this conclusion including several lags and get very low elasticities. Meanwhile, Jimeno and Bentolila (1998) point that migration decisions are poorly sensitive to the unemployment rate and real wages. In line with these results, Antolin and Bover (1997) conclude that unemployment rate has not a significant effect in the international migration. However, this apparent puzzle is observed not only for Spain, being that Italian internal migrations do not react to mass unemployment, as Fachin (2007) pointed out. Antolin and Bover (1997) also show that emigration is produced from regions where wages are higher than the average in the period 1987-1991, which opposes to the theoretical results.

Figure 1 shows the evolution of the gross internal migration and the unemployment rate in the extended period 1988-2010. The figure displays that the lower is the unemployment rate, the higher is the interregional gross migration rate in Spain.

Figure 1. Gross interregional migration and unemployment rate, 1988-2010.



Source: EVR, INE.

These counterintuitive results, and against the hypothesis considered in the majority of the theoretical models, have been analyzed with different techniques. Maza and Villaverde (2004) use a semi parametric model, capturing a weak effect of unemployment rate on internal migration for the period 1995-2002. Juarez (2000) considers the inflows and outflows for the period 1962 to 1993 instead of net migration flows. This paper analyze the wide definition of migration where pull factors of receiving region and push factors of origin regions are included in the model. He includes a non-linear effect in the unemployment rate and concludes that, under a certain specification, gross migration flows respond with the expected sign to both unemployment rates and wages differentials. Similarly, Muhler and Watson (2010) also obtain appropriate signs for the coefficients of unemployment and wages. They consider inflows vs. outflows for the period 1990 to 2000 and allow for the presence of some structural breaks across this sample. These authors demonstrate that the inclusion of housing prices in the empirical study is important in order to obtain the expected effect of unemployment and wages, and they conclude that the effect of labor market variables is different from the period 1990-1995 to the period 1996-2000. However, the period of change is exogenously selected.

The aim of this paper is to determinate the effect of the labor market variables in the Spanish internal migrations performed in the period 1988-2010. By extending the Harris and Todaro (1970) model, where expected wage is considered, a simple theoretical framework is developed to demonstrate that, instead of a structural break in a given period, the migration reacts differently depending on the level of the expected wage. This enhances us to empirically test the implications of the model. A model with an endogenously method of selection of the threshold value is considered in the econometric specification and, by following Juarez (2000), migration is defined for every region as an outflow from origin region to destination region, therefore, each data indicates the number of outflows from region i to region j. Furthermore, our perspective is that the regional flows is not only caused by labor market variables, because some elements associated to the regional landscape could be key factors that could explain the observed behavior of regional mobility. As the current economic crisis shows, the housing price could be a good proxy of migration costs and it has been taken into account in the migration process.

In this paper it is showed that, when the labor market is characterized by better labor conditions in the destination region, the decision to migrate responds to labor market variables as the expected wages. By contrast, when the distance between labor markets fundamentals is advantageous for the source region, labor market factors are less significant, if any, for the migration decision.

This paper is organized as follows. Section 2 presents the theoretical model on which we base our empirical exercise. Section 3 presents some stylized facts and describes the data employed and the econometric methodology. Then, in Section 4, we show the empirical results obtained and, finally, in Section 5 we summarize the main conclusions.

2 Theoretical model

This Section is devoted to develop a basic theoretical framework. The model consists of an economy with two locations, region of origin (*i*) and host region (*j*). Every individual lives one period and, at the beginning of the period, he faces the possibility of migrating to another region different from birth region. This decision is taken by comparing the utility derived in both places. Therefore, the total emigration from the origin region has the following form:

$$m_{i,j} = f\left(\frac{U_j}{U_i}\right)$$

where $\frac{\partial m}{\partial U_j} > 0$ and $\frac{\partial m}{\partial U_i} < 0$.

Let us consider an individual deriving utility from goods consumption *c* and the housing good h in the residence region r. In particular, we assume the following utility function, which takes a standard Cobb-Douglas form:

$$U_r = c_r^{\propto} h_r^{\beta} \tag{1}$$

with r=i,j, and $\propto, \beta > 0$. The parameter $\propto$ reflects elasticity of the consumption good and parameter β reflects the elasticity of the housing good.

Every worker has to contribute to an unemployment insurance fund and this is assigned to unemployed people. Therefore, the perceived wage has to be discounted in the proportion given to non-employed workers, that is, wage multiplied by one minus the unemployment rate.

The budget constraint considers the consumption good price as a constant equal to one in the region of birth and the host region. The price of the houses, *p*, is a relative price that is not assumed constant and it is different in the origin and host region. The positive relationship between housing prices and wages is already presented in Carliner (1973), among others. Therefore, it is assumed that housing prices depend positively on the perceived wage in both places, origin and destination, because the houses are not only bought by residents.

Consequently, the budget constrain is the following:

$$c_r + p_r(w_r^e, w_{-r}^e)h_r = w_r^e \tag{2}$$

with $\frac{\partial p_r(w_r^e, w_{-r}^e)}{\partial w_r^e} > 0$ and $\frac{\partial p_r(w_r^e, w_{-r}^e)}{\partial w_{-r}^e} > 0$, being *r* and –*r*, the residential and the non residential areas respectively, and w_r^e the expected wage in each area (that is, the wage weighted by unemployment rate).

Solving the problem of maximizating equation (1) subject to the budget constraint (2), the optimal consumption and housing goods are characterized by the following values, respectively:

$$c_r = \frac{\propto w_r^e}{\propto + \beta} \tag{3}$$

$$h_r = \frac{\beta w_r^e}{(\propto + \beta) p_r(w_r^e, w_{-r}^e)} \tag{4}$$

From equations (3) and (4) it is straightforward to check that a higher expected wage in the origin region increases the purchase consumption goods, but the effect is indeterminate in the case of the purchase of housing good.

The effect of an increase of the expected wage in the region -*r* in the utility of the region *r* is negative, while the effect on an increase in the expected wage in region *r* in the utility of region *r* is also negative if

$$\frac{w_r^e}{p_r(w_r^e,w_{-r}^e)}\frac{\partial p_r(w_r^e,w_{-r}^e)}{\partial w_r^e} > \frac{\propto+\beta}{\alpha} \qquad (5)$$

That means, if the elasticity of the housing price to the expected wage is high enough, higher than one, what occurs in the goods considered as luxury, the effect of an increase in the wages could increase so much the housing price that the utility on the region could decrease.

Inequality (5) will fulfil the greater are the expected wages in the residence region, in this case, the migration does not react conventionally to an increase in the expected wages. If equation (5) accomplishes, an increase in the income perceived in the region of origin could act as push factor for emigration.

Summarizing, when the labor market is characterized by low expected wages in the origin country, the decision of migrating responds to labor market variables. By contrast, when the labor market conditions in origin are high enough, labor market factors are less significant, and could even react opposite to the results derived from neoclassical migration models. The model also highlights that other institutional factors as housing prices are also explanatory elements. Next section will test the main implications of the model.

3 Data and Empirical Methodology

The statistical sources of our annual frequency database, which covers the period 1988-2010 for the 17 Spanish regions or Autonomous Communities, are captured from different databases. Data of the unemployment rate and the labor force of each region has been obtained from the working population survey (Encuesta de Población Activa, EPA) provided by Spanish National Statistics Institute (Instituto Nacional de Estadística, INE). Nominal wages are calculated by dividing total employees earnings by the number of employeed workers, both series obtained from the Statistics of Spanish Ministry of Economy and Competiveness. This data has been deflated by CPI index in order to obtain real wages. Housing prices data is provided by the Spanish real estate valuation society "Sociedad de Tasación", the only source of information that covers the analyzed period and disaggregates across regions. Finally, geographical distance is measured as the kilometers between capitals of each region. The series that have a higher frequency have been annualized by taking the annual average value.

A first approach to data shows a non-conventional influence of some labor market conditions on migration rates in Spain. The first difference is the counterintuitive evolution of the variables unemployment and emigration presented in Figure 1. Another relevant characteristic is the different influence of wages and unemployment dispersion on migration illustrated by Figure 2 and Figure 3. We have considered the variation coefficient between regions as measure of wage and expected wage dispersion.

Figure 2 does not show a clear influence of wage dispersion on migration. Moreover, the relationship between migration and regional unemployment dispersion, Figure 3, is null, and it is one of the reasons for the so-called Spanish puzzle in internal migration.

Figure 2. Regional wage dispersion in Spain and gross migration rate: 1988-2010

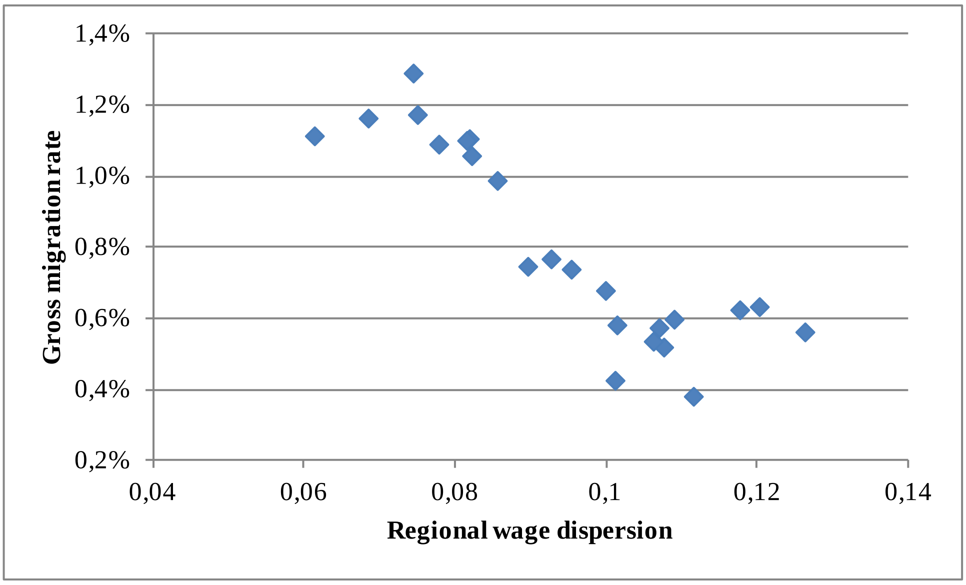
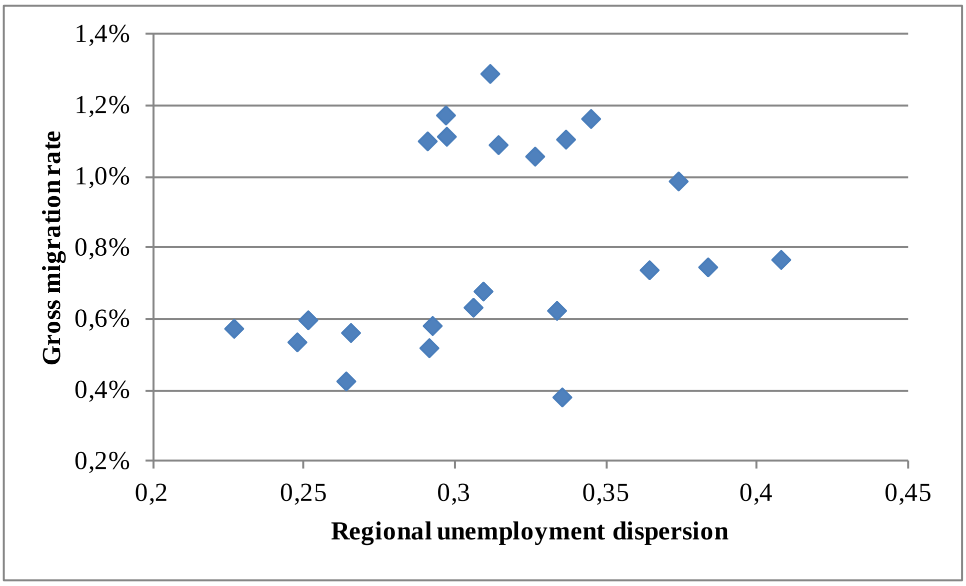


Figure 3. Regional unemployment dispersion in Spain and gross migration rate: 1988-2010.



Therefore, these special characteristics of interregional migration in Spain lead us to think that the labor market variables chosen in the empirical literature may not be correct because the labor market conditions could not be represented by isolated variables as wages and unemployment.

Therefore, our perspective is that neither the wages nor the unemployment rate explain migration process themselves, but is the combination of both factors which migrants take into account when moving. Thus, we will include the expected real wage as key variable, which takes into account both the potential wage earnings and the probability of finding a job. Despite the fact that the expected wages have been considered in many theoretical papers to analyze migrations, empirical papers do not usually capture this idea.

Our estimation procedure has two steps. First, a simple regression with real wages and unemployment as independent variables is made to justify the usefulness of including the aggregate variable expected wages. Apart from using the labor market conditions and housing prices to explain the migrations, some other variables have been considered. Based on the gravity equation, which has been applied to the study of migration flows since Greenwood (1975), migrants make their decisions according to two sets of variables. On the one hand, the attractive forces of the destination and, on the other hand, the costs of migration. If the expected value of the first factors, where we can include economic, social or institutional factors, outweighs the latter, the migration will take place.

Assuming, as usual in the related literature, that the labor force size in the origin and in the destination regions contribute positively to migration, we will consider as attractive forces higher expected wage in the target region and lower housing prices in the origin region. The costs of moving from the origin place to the destination would be the distance between regions, which also is a proxy of other costs of migration, the housing prices in the receiver region and lower expected real wage in the origin. Against this background, we can initially pose the Model I:

$$\ln M_{ij,t} = \beta_0 + \beta_1 \ln n_{i,t}$$
$$+ \beta_2 \ln n_{j,t} + \beta_3 \ln h_{i,t} + \beta_4 \ln h_{j,t} + \beta_5 u_i + \beta_6 u_j + \beta_7 \ln \mathrm{w}_{\mathrm{i,t}} + \beta_8 \ln \mathrm{w}_{\mathrm{j,t}}$$
$$+ \beta_9 \ln d_{ij} + \beta_{10} origin + \varepsilon_{ij,t}$$

In this specification, M_{ij} represents the migration flow, that is, the number of people who migrates from region i to region j, n is the labor force, h is the housing price, u is the unemployment rate, w is the real wage, d_{ij} is the geographic distance between region i and j and ε_t represents the perturbation of the model. We have additionally included 16 dummy variables in order to capture the idiosyncratic effect of each of the source regions, which take value 1 for the origin region and 0 otherwise. For purposes of parsimony, we will not include other kind of fixed effects as dummies by the destination region or by year.

In order to test whether the model specification is improved when the expected wage variable is introduced instead unemployment and wages separately, we will propose the Model II:

$$\ln M_{ij,t} = \beta_0 + \beta_1 \ln n_{i,t} + \beta_2 \ln n_{j,t} + \beta_3 \ln h_{i,t} + \beta_4 \ln h_{j,t} + \beta_5 \ln w^e_{i,t} + \beta_6 \ln w^e_{j,t}$$
$$+ \beta_7 \ln d_{ij} + \beta_8 origin + \varepsilon_{ij,t}$$

where $w^e_{ij,t} = w_{ij,t}(1 - u_{ij,t})$ is the expected real wage.

It would seem reasonable that, when the relative opportunities in the labor market of the destination region are higher, the decisions of migrants fit traditional patterns whilst, if the economic gain of moving is relatively low, migrants attend also others factors, becoming the labor market variables less decisive. Therefore, the model can be nonlinear. In order to capture this nonlinearity, in a second step a model with an endogenously method of selection of the thresholds is considered. So, we have re-specified Model III as follows:

$$\ln(M_{ij,t}) = z_{ij,t}\delta_1 + \epsilon_{1,ij,t} \quad \text{if} \quad q_{ij,t} \leq \gamma$$

$$\ln(M_{ij,t}) = z_{ij,t}\delta_2 + \epsilon_{2,ij,t} \quad \text{if} \quad q_{ij,t} > \gamma$$

where $z_{ij,t} = [1, ln(n_{i,t}, n_{j,t}, h_{i,t}, h_{j,t}, w^e_{i,t}, w^e_{j,t}, d_{ij}), origin\,]$, 1 reflects the intercept, δ_1 and δ_2 are two vectors of parameters and $q_{ij,t}$ is the threshold variable. If the parameter γ were known, we could easily test for the nonlinearity null hypothesis and estimate the system. However, this parameter is unknown and, consequently, we should first estimate it. To do this, we have carried out a grid search, using a 15% trimming. The estimated value of the parameter γ coincides with the value that maximizes the LM statistic for testing the $\delta_1 = \delta_2$ null hypothesis. In order to test for this hypothesis, we have additionally calculated the critical values of the distribution by way of bootstrap techniques. In order to ensure that the critical values provide sufficient reliability, we have conducted 500 replications, obtaining the critical values of X and Y for a 5% and a 10% significance level, respectively.

4 Results

In this Section we detail the econometric outcome obtained. We show the results of the estimation of the Models I and II with thresholds in Table 1, and the model III in Table 2, in which we present the elasticity coefficients of each variable. We should note that the t-ratios are corrected by the White statistic, taking into account the presence of heteroscedasticity.

In the Model I exposed in Table 1, It can be checked that the unemployment rate in origin is no significant in explaining the interregional migrations in Spain and that the unemployment rate in host region has the opposite sign, that is, if the unemployment rate in destination increase a 1%, the immigration to this region will increase a 0.07 %. This has been corrected in Model II by introducing expected wages as an aggregate variable, and all signs obtained are the expected. Total working population in both origin and destination regions affect positively to the migration flow. Housing prices in the source region also have a positive relationship with the migration flow, since migrants try to minimize that cost. By contrast, higher housing prices in the region of destination discourage the migration decision. Finally, geographic distance influences negatively migration flows. Model II let us to conclude that a 1% of increase in the expected wage in origin decrease the emigration from this region 0.23%, while the pull factor of the host region is greater.

Regarding the Model III, displayed in Table 2, the upper part corresponds to the estimation of the value of the threshold γ, the value of the LM statistic, which contrasts the null hypothesis of no presence of threshold, and the joint model R^2, while second and third part exhibits the elasticities of the different models for $q_{ij,t} \leq \gamma$ and $q_{ij,t} > \gamma$ respectively.

By following the results derived in the theoretical framework a threshold in expected wages is considered, but as this variable is built with the regional wages and the unemployment rates, both variables are also considered separately as threshold. Therefore, we consider three labor market gaps as threshold variables in order to contemplate different alternatives. We estimate γ and, then, we test the null hypothesis of no presence of threshold. In all cases we can reject the null hypothesis, so we can confirm the existence of different behaviors depending on the value taken by the threshold variable considered.

We begin by considering the case of the unemployment rate as the threshold variable. As seen in the last column of Table 2, the migration behavior in the two different parts of the sample, that is, for smaller and larger values of u_i/u_j=0.81, is different and all the coefficients are significant and have the expected sign. When the unemployment rate in the source region is lower in a certain proportion than the rate in the target region, the importance of the size of labor force in both regions is lower. By contrast, the housing price and the expected real wage in the source region are more decisive. Geographical distance is not very influent, as well as real estate prices and expected real wages in the destination region.

We now analyze the estimation of the model when we allow for thresholds depending on the real wage gap. We should note that the interpretation in this case is contrary, because sample for $q_{ij,t} \leq \gamma$ refers to a worse labor conditions in the origin region. The value of the threshold for which migration movements change the response to the explanatory variables is 0.92, a ratio greater than the previous case. In this case, all the variables included are more determinant when the labor conditions are not favorable to the origin region and the power of explanation of the model is higher. In fact, expected real wages in the source region are not significant when the labor conditions are advantageous.

Finally, the gap in expected real wages as threshold variable is considered. This case is similar to the previous one in terms of interpretation, with a very close threshold value and the lack of significance of some coefficients. The size of the labor force of both regions is more determinant when the expected wage in origin is less than 92.1% of the expected wage in the host region and also the coefficient of explanation is higher. The same happens with the housing price of destination and the distance, which affect negatively the outflows. The greatest difference is that labor market conditions in origin or destination do not affect migrations when the expected wage in origin is high enough.

From Table 2, it can also be concluded that the push factors of origin region have a higher effect on migration than the pull factors of hosts regions.

5 Conclusions

In this paper we have analyzed the interregional migrations in Spain during the period 1988-2010 by considering origin and destination labor market conditions. A simple theoretical model is developed to capture the idea that migration reacts differently to different levels of expected wages instead reacting differently to structural breaks. The implications of the model are tested empirically. Our perspective is innovative because nonlinearities in migration flows are allowed by including thresholds in the labor market conditions. Results indicate that, when labor market conditions are unfavorable to the origin region, migrants make their decisions based on the labor market conditions and housing price in origin and destination as well as on distance between the regions. All the variables have a high capacity of explanation, being greater those that capture the push factors instead the pull factors. However, when relative labor market conditions in the source region are good enough, migration flows are less responsive, or even indifferent, to expected wages, suggesting that migrants are less sensitive to labor variables.

References

Antonlin, P. and Bover O. (1997) “Regional Migration in Spain: The Effect of Personal Characteristics and of Unemployment, Wage and of Housing Price Differentials Using Pooled Cross-Sections,” *Oxford Bulletin of Economics and Statistics*, 59, pp. .

Bentolila, S. and Dolado, J.J. (1990) “Mismatch and Internal Migration in Spain,” in F.Padoa-Schioppa, ed, Mismatch and Labor Mobility, Cambridge University Press.

Bentolila, S. (2001) “Las migraciones interiores en España,” Documento de trabajo FEDEA nº 2001-07.

Carliner, G. (1973): "Income elasticity of housing demand", *The Review of Economics and Statistics*, 55(4), 528-532.

Fachin, S. (2007) "Long-run trends in internal migrations in Italy: a study in ppanel cointegration with dependent units" *Journal of Applied Econometrics*, 222, 401-428.

Greenwood, M. (1975): “Research on Internal Migration in the United States: a Survey,” *Journal of Economic Literature*, 13, pp. 397-433.

Harris, J.and Todaro,M. (1970) “Migration, Unemployment and Development: A Two-Sector Analysis,” *American Economic Review*, , pp. 126-142.

Jimeno-Serrano J.F. and Bentolila, S. (1998) "Regional Unemployment Persistence (Spain 1976-94)," *Labour Economics*, 5, pp. 25-51.

Juarez, J. P. (2000): “Analysis of Interregional Labor Migration in Spain Using Gross Flows,” *Journal of Regional Science*, 40(2), pp. 377-399.

Maza, A. and Villaverde, J. (2004) “Interregional Migration in Spain: A Semiparametric Analysis,” *The Review of Regional Studies*, 34, pp. .

Mulhern, A. and Watson, J. (2012) “Spanish Inter-Regional Migration: an Enigma Resolved,” *Applied Economics Letters*, 17, pp. 1355-1359.

Table 1: Models without Thresholds

	Model I	Model II
Intercept	-5.99	-4.07
Labor force$_i$	1.3	1.19
Labor force$_j$	0.94	0.94
Housing price$_i$	0.51	0.36
Housing price$_j$	-0.53	-0.41
Unemployment$_i$	0.06**	
Unemployment$_j$	0.07	
Wage$_i$	-0.81	
Wage$_j$	1.06	
Expected wage$_i$		-0.23
Expected wage$_j$		0.39
Distance$_{ij}$	-0.65	-0.63
R^2	0.76	0.76
Observations	6256	6256

* and ** denotes no significance at 5% and 10 % respectively.

Table 2: Model III with thresholds

	q: w_i / w_j	*q*: $w^e_{i,t} / w^e_{j,t}$	*q*: u_i / u_j
*q**	0.921	0.92	0.813
LM Test	200.75	216.25	237.99
R^2	0.83	0.83	0.83
	q<*q**		
Intercept	-1.94	-1.88**	-4.93
Labor force$_i$	1.75	1.81	0.79
Labor force$_j$	1.12	1.18	0.93
Housing price$_i$	0.44	0.77	0.59
Housing price$_j$	-0.89	-1.02	-0.31
Expected wage$_i$	-0.76	-1.06	-0.45
Expected wage$_j$	0.90	0.83	0.36
Distance$_{ij}$	-0.73	-0.61	-0.38
R^2	0.83	0.84	0.84
	q>*q**		
Intercept	-5.43	-5.75	-2.67
Labor force$_i$	0.99	1.05	1.40
Labor force$_j$	0.86	0.87	0.97
Housing price$_i$	0.44	0.25	0.22
Housing price$_j$	-0.35	-0.34	-0.51
Expected wage$_i$	0.007*	0.19*	-0.31
Expected wage$_j$	0.23**	0.25	0.60
Distance$_{ij}$	-0.56	-0.60	-0.77
R^2	0.81	0.73	0.71

* and ** denotes no significativiy at 5% and 10 % respectivelly.