On the geography of unemployment rates and the spatial sorting of workers' schooling

(Preliminary version prepared for the 2013 Applied Economics Meeting. In case of interest request the authors an updated version)

Enrique López-Bazo [1]
Elisabet Motellón [1, 2]

1 AQR—IREA. Universitat de Barcelona
Avda Diagonal 690. 08034 Barcelona. Spain
Tel: +34 93 4037041 FAX: +34 93 4021821

2 Universitat Oberta de Catalunya
Avda. Tibidabo 39-43. 08035 Barcelona. Spain
Tel: + 34 93 2537501

Email: elopez@ub.edu; emotellon@uoc.edu

Abstract: A distinctive feature of unemployment is that its incidence is far from being homogeneously distributed in the territory. Disparities in unemployment rates are not only observed between countries but also between regions within countries. The available evidence indicates that since the early 80s Spain is a country of high unemployment rates, and persistently large regional disparities. In fact, there is a clear spatial divide in regions showing higher than the average rates in the Northeast, and regions with rates below the average in just a few hundred kilometres distance, in the Southwest. In contrast with the previous studies, here we use micro-data for the Spanish NUTS3 regions to explore the relationship between the spatial distribution of individuals' education and regional unemployment rates. We provide novel evidence showing that i) the impact of individual's education on unemployment largely varies across regions and, ii) regional disparities in the level of educational attainment of the active population explain a big deal of the observed disparities in the regional distribution of unemployment rates, particularly in periods of high unemployment.

JEL codes: C25, J64, R23

Keywords: Regional labour markets, Regional unemployment, Education, Spatial dependence

Acknowledgments: The authors wish to acknowledge financial support from the Spanish Ministry of Economics and Competitiveness, National Program of R&D, ECO2011-30260-C03-03, and the European Community's Seventh Framework Programme (FP7-SSH-2010-2.2-1) under grant agreement n° 266834, SEARCH Project.

1. INTRODUCTION

Unemployment is one of the issues in the agenda of academics and policy makers due to its obvious social and economic implications. This is even more so in periods of recession, particularly in economies with a larger share of the active population unemployed. Another distinctive feature of unemployment is that its incidence is far from being homogeneously distributed in the territory. Disparities in unemployment rates are not only observed between countries but also between regions within countries. Interestingly, such regional disparities persist over time, which means that there are territories in a country showing unemployment rates well above those of some other regions for decades, with no churning at all or even without a trend to converge to similar rates.

The available evidence indicates that since the early 80s Spain is a country of high unemployment rates, and persistently large regional disparities. In fact, there is a clear spatial divide in regions showing higher than the average rates in the Northeast, and regions with rates below the average in just a few hundred kilometres distance, in the Southwest. This regional gap evolves with the business cycle. In the most recent period, it decreased in absolute terms during the boom of the Spanish economy from the late 90's to the beginning of the crisis at the end of 2008, and it increased suddenly and steadily from then on. This change caused by the current crisis, and the existence of spatial differences in the incidence of the housing bubble and the ulterior impact on the building and banking sector, makes the Spanish regions be an interesting case study.

As indicated in the literature review of the next section, previous contributions have studied regional unemployment in Spain from different angles, though most of them adopting an aggregate or macro approach. That is to say, they have tried to explain the features in the regional distribution of unemployment rates by means of macro factors and exploiting aggregate regional data. A similar approach has been followed in studies dealing with regional unemployment in other countries. Actually, the existing evidence (see for instance OECD, 2011a) confirms that Spain is not a *black swan*, in the sense that sizeable regional disparities in unemployment rates are observed in a large number of developed countries. In any case, the only study that we are aware of that depart from the aggregate approach is López-Bazo and Motellón (2012). Using the micro-data from the Labour Force Survey (LFS) in each NUTS2 Spanish region, they assessed the contribution of a set of individual characteristics in explaining the regional gap in unemployment rates.

In this study, we take the route proposed in López-Bazo and Motellón (2012), using the LFS micro-data of the Spanish NUTS3 regions to explore the relationship between the spatial distribution of individuals' education and regional unemployment rates. More concretely, we aim at checking if regional disparities in the level of educational attainment of the active population explain most of the observed disparities in the regional distribution of unemployment rates. To do so, we apply standard tools for describing the spatial distribution of unemployment rates, and a probit model for estimating the impact of individual's schooling in each region, controlling for the likely sample selection caused by the individual's decision to participate or not in the labour market. As a novel contribution to the literature, this procedure allows us to describe the spatial pattern of the impact of education on the unemployment rate, and to compute counterfactual rates for each province under different scenarios regarding the level of schooling. The comparison of the actual and counterfactual distributions of regional unemployment rates is used to conclude on the effect of the spatial sorting of workers' education on the geography of unemployment rates in Spain.

The rest of the paper is organised as follows. A brief literature review on the analysis of regional unemployment is provided in section 2. Section 3 presents the dataset and comments on the main variables used in the analysis. It also discusses the periods and the sets of regions under analysis. The description of the regional distribution of unemployment rates in Spain in the periods under analysis, and that for the measure used to proxy individual's education, is provided in section 4. Section 5 introduces the empirical model used to estimate the impact of education on the individual's probability of unemployment in each region, while section 6 discusses the results. Finally, section 7 concludes.

2. BRIEF LITERATURE REVIEW

Disparities in unemployment rates are sizeable and persistent both between and within many countries (OECD, 1989, 1990, 2000, 2005; Blanchard and Katz, 1992; Decressin and Fatas, 1995; López-Bazo et al., 2002; Overman and Puga, 2002; Cracolici et al., 2007; Bande et al., 2008; Filiztekin, 2009). In Spain, for example, data from the Labour Force Survey (LFS) reveal that the unemployment rate in 2010 was 20.2%, the highest among OECD members. That year, the unemployment rate in Andalusia and the Canary Islands reached 28% and 28.7% respectively, whereas at a distance of a few hundred kilometres and within a similar institutional framework, the rate in the Basque Country was 10.5%.

As such, the regional gap in Spain's unemployment rate is of a similar order of magnitude to that observed between the country considered as a whole and rates recorded in EU and OECD economies. Furthermore, the analysis in OECD (2005) suggests that Spain is no anomaly here, as the degree of regional disparity in unemployment rates within a number of countries (including Germany, Italy, Mexico, and Turkey) is even higher than that observed in Spain. These results also point to the increase in regional inequality within countries as being at the root of the intensification of overall inequality in unemployment rates in Europe.

From a theoretical perspective, in a world characterized by the absence of adjustment costs and rigidities, disparities in unemployment rates across locations would not be expected to persist. Situations of excess labour in one area would quickly disappear as workers moved to areas with higher rates of unemployment. However, the evidence (Lazar, 1997; Evans and McCormick, 1994; Martin, 1997; Martin and Sunley, 1999; Overman and Puga, 1999; López-Bazo et al., 2005) indicates just the opposite: regions with high unemployment in a given decade continue to suffer high unemployment rates in the following decades, while regions with low unemployment continue to enjoy low rates.

The slow wage adjustment rate and the high costs incurred by individuals and firms when migrating probably explain why idiosyncratic shocks, or contrasting regional responses to common shocks, might cause unemployment rates to differ markedly across regions for long periods. Given this explanation, heterogeneity in the spatial distribution of unemployment can be seen as what Marston (1985) defines as a disequilibrium phenomenon. A second explanation as to why certain areas have differing unemployment rates is also provided in Marston (1985), drawing on ideas in Hall (1972) and Rosen (1974). A steady-state relationship in unemployment rates across regions exists as a function of their factor endowment and since this endowment differs from one region to another, the spatial distribution of unemployment is not homogeneous. Moreover, as long as this endowment remains stable, the distribution of unemployment will not change dramatically. This equilibrium hypothesis, therefore, is based on the idea that workers have incentives not to migrate when unemployed because, for one reason or another, they value these endowments. On the other hand, when selecting their optimal location, firms take into account other regional endowments in addition to those of wage and unemployment rates (Partridge and Rickman, 1997). Evidence regarding high wages in areas of high unemployment supports this view, as does the preference for certain facilities and amenities. Martin (1997) and Partridge and Rickman (1997) extend the list of factors that might account for unemployment equilibrium differentials to permanent differences in economic, institutional and labour market characteristics across regions.

Most previous contributions to the empirical literature (Elhorst, 1995; Partridge and Rickman, 1997; Taylor and Bradley, 1997; López-Bazo et al., 2002, 2005; Filiztekin, 2009; Bande and Karanassou, 2009) have sought to analyse the determinants of regional inequalities in unemployment by means of an aggregate specification in which the unemployment rate in each region, or the deviation from a benchmark (the nationwide average or the region with the lowest rate), is related to regional magnitudes proxying for both the disequilibrium and the equilibrium determinants of unemployment. It should be noted that this aggregate approach imposes the same effect on each variable in all regions, while only partially (and thus imperfectly) accounting for regional heterogeneity in individual and household characteristics, i.e., for the sorting of individuals across regions according to their observed characteristics.

The expected impact of education on unemployment can be used to illustrate our argument. A rising level of education in a region is assumed to have a negative impact on its rate of unemployment, given that findings at the micro level suggest that education increases the probability of an individual finding and keeping work (e.g., Mincer, 1991; OECD, 2011b). Accordingly, the effect of the regional endowment of education on the regional unemployment rate is estimated to be negative and significant in six out of the nine studies reviewed by Elhorst (2003). Yet, contradictory findings are reported in the remaining three studies. Furthermore, the effect is reported as being positive, and in some cases even significant, for the set of Canadian regions in Partridge (2001), while no significance was found for the Spanish regions in López-Bazo et al. (2002, 2005). Likewise, Filiztekin (2009) finds no evidence of a robust negative effect for the Turkish provinces. Thus, there would seem to be some contradiction between the expected effect of education on an individual's probability of unemployment and the findings of empirical studies using micro-data, and (at least part of) those of aggregate studies using regional data.

3. DATA

Almost all previous studies in the literature on regional unemployment draw on statistical information at the aggregate regional level. To our knowledge, the only exception so far is López-Bazo and Motellón (2012), in which we exploited the micro level data contained in the Spanish wave of the Labour Force Survey to analyse the contribution of individuals' characteristics on the regional gap in unemployment rates between the Spanish NUTS 2 regions. In a similar vein, this study is based on individual data for each of the Spanish provinces.

The data correspond to the second quarter for years 2006, 2007, 2011, and 2012, contained in the LFS.[1] The LFS is produced by the Spanish National Institute for Statistics in line with the criteria laid down by EUROSTAT for EU Member States. The survey provides information about the status of individuals in the labor market (non-participant, employed, unemployed) and the characteristics of individuals and households (gender, nationality, age, education, number of household members, etc.). The sample used in our analysis comprises individuals between 16 and 65 years of age in each of the fifty NUTS 3 regions in Spain (provinces). Provinces have been frequently referred as to the functional labour market areas in Spain, due to the fact that most of commuting flows are observed within the province boundaries. Notice that the LFS-sample design ensures that it is representative of each of the Spanish provinces.

[1] The LFS is conducted each quarter. However, given the impact of seasonality on Spain's labour market (being particularly sensitive to tourism and other activities in the service and primary sectors), we present the results using information for the second quarter of each year, as it would seem that this quarter is the one that is least influenced by seasonal variations. Note, nevertheless, that our results are robust to the consideration of data for the other three quarters.

In order to distinguish between the effect of individuals' schooling on the unemployment rate in each province in periods of economic boom (lowest unemployment) and during recessions (highest unemployment), we focus the analysis in two periods. The first one, comprising the years 2006 and 2007 is the one in which the figures of unemployment in the Spanish economy (and in each of the provinces) were the lowest in decades. In the country as a whole, the rate of unemployment was 8.53% in the second quarter of 2006, and 7.95 in the same quarter of 2007. From that moment on, unemployment started growing rather rapidly. The second period in our analysis is the one corresponding to the current crisis, 2011 and 2012. In the second quarter of 2011 the unemployment rate had increased up to 20.89%, and in the same quarter of 2012 the Spanish unemployment rate climbed to 24.63%.

It needs to be said that our analysis in the following sections pools the information contained in the second quarter of the LFS in each pair of years to mitigate the impact of peculiarities associated to each one of the years, and most importantly, to increase the number of available observations for some of the smaller provinces. Undoubtedly, the cost of such decision is to impose homogeneity in the impact of schooling, and the other factors included in the model aiming at explaining unemployment in each province, in each of the two subsequent years for each period under analysis. In any case, it must be stressed that the results obtained when using observations for one of the couple of years used in each of the periods led to similar conclusions to the ones obtained with those showed in the following sections.

4. DESCRIPTIVE ANALYSIS

The unemployment rates in each of the provinces in the periods of high, 2011-12, and low unemployment, 2006-07, are shown in the first two columns of Table A1 in the Appendix. To easy the interpretation of this information, we have estimated non-parametrically, using the kernel method, the density function for the unemployment rates in each of the periods, and plot it in Figure 1. It is observed that in the last period unemployment rates were not only much higher than before the current crisis, but that dispersion in the provincial distribution was also much more pronounced. Actually, the distribution corresponding to 2006-07 is fairly concentrated around the mode, with only a remarkable mass of probability in the right tail of the distribution, on rates of unemployment around 12-14%. On the contrary, the 2011-12 distribution is much flatter, indicating that the labour market reacted differently to the deterioration of the economy in each province, causing huge regional disparities in unemployment rates. In this period, there are provinces with unemployment rates between 10% and 15%, and others with figures around, and even above, 30%.

A simple inspection of the spatial distribution of unemployment rates in Figure 3 confirms what has already been reported elsewhere: there exists a sort of North East-South West divide in the incidence of unemployment in Spain. The choropleth maps in Figure 3 show the quantile in which each province is placed depending on its unemployment rate in each period. The divide seems to be clearer in 2006-07 than in 2011-12, mostly due to the higher rates in the Mediterranean (coastal East) provinces, and the lower relative incidence of unemployment in the North-western provinces. In that regard, it must be noticed the effect of the larger incidence of the building sector in the Mediterranean provinces during the period of expansion and housing bubble, and the stronger impact on job destruction in that area following the bubble burst.

In any case, the maps suggest strong spatial dependence in the provincial distribution of unemployment rates in Spain, that are confirmed by results obtained for the so-called I-Moran statistic reported in the first group of rows of Table 1. Both in the case of a binary contiguity and of an inverse squared distance matrix, the null of absence of spatial dependence in unemployment rates in the set of Spanish provinces is strongly rejected. These results thus agree with previous evidence reported for Spain and other economies worldwide (e.g. Filiztekin, 2009; Patacchini and Zenou, 2007; López-Bazo et al., 2002).

In this section we also provide some descriptive evidence regarding the provincial distribution of the average years of schooling of the active population. Our motivation here is simply to show that the provincial distribution of schooling is far from homogeneous, that it is not spatially random, and that it might be connected to the pattern observed for unemployment rates. The figures on the years of schooling in each province are reproduced in the last two columns of Table A1 in the Appendix. Such figures were used to estimate the density functions for the two periods under analysis in Figure 2, and to produce the choropleths maps in Figure 4.

The estimated densities reveal the shift to the right (towards higher values) of the distribution of years of schooling and, even more interestingly for our analysis, certain increase in the amount of dispersion. The increase in dispersion seems to be caused by a fatter right tail in 2011-12. Actually, whereas the mass of probability at very low levels of schooling in 2006-07 vanish in the distribution corresponding to 2011-12, that associated to (relative) mid-high values of schooling persists and even increases. In any case, the classification of provinces in quantiles of the schooling distribution depicted in the maps of Figure 4 suggests the existence of a spatial divide, with higher educated individuals in the North and less educated ones in the South. Although there are a few exceptions (such as Seville in the South), the comparison of these two maps with the ones corresponding to the spatial distribution of unemployment in Figure 2 support the hypothesis that individuals' schooling is likely to determine their chances to be unemployed, and thus to explain part of the disparities in unemployment rates observed across provinces. As a final piece of descriptive evidence, results of the I-Moran test reproduced in Table 1 confirm the existence of strong spatial dependence in the provincial distribution of years of schooling.

In the following sections we explore in detail the connection between schooling of the active population and the probability of being unemployed in each province, and assess the impact of the spatial sorting of education on the observed provincial disparities in unemployment rates.

5. EMPIRICAL MODEL & STRATEGY

The empirical setting for the assessment of the impact of schooling and the other observed characteristics assumes that there is a latent equation linking the probability of an individual i in province p being unemployed to an individual's set of characteristics:

$$\mathrm{prob}(U)_i^p = \mathbf{X}_i^p \boldsymbol{\beta}^p + \varepsilon_i^p \tag{1}$$

where prob(U) denotes the probability of unemployment, $\mathbf{X}$ includes the aforementioned characteristics, $\boldsymbol{\beta}$ is the corresponding vector of coefficients, and $\boldsymbol{\varepsilon}$ is an error term.[2]

However, our empirical setting assumes that, in a first step, all individuals in each province face the decision to participate or to not participate in the labour market. That is to say, given the situation of the labour market and the own conditions and characteristics, individual i in province p chooses being active or being out of the labour market. At a latter stage, those individuals that decide to participate are then classified as employed or as unemployed depending on their success in occupying a job. Formally speaking, we assume there is a latent relationship for the probability of participating in the labour market such as:

$$\mathrm{prob}(P)_i^p = \mathbf{Z}_i^p \boldsymbol{\gamma}^p + \mathrm{v}_i^p \tag{2}$$

where prob(P) denotes the probability of being active, $\mathbf{Z}$ includes a set of individual and household characteristics, $\boldsymbol{\gamma}$ is the corresponding vector of coefficients, and $\boldsymbol{\nu}$ is an error term.

The estimate of the impact of schooling on the probability of unemployment based just on eq. (1) will be appropriate only in case there are no systematic differences in the sample of individuals participating and non participating in the labour market. Otherwise, a sample selection issue will bias the estimates, causing misleading conclusions on the impact of education on unemployment. In such a case, consistent estimates can be obtained by mean of the sample selection probit model, suggested by Van de Ven and Van Pragg (1981).

In a nutshell, we can only observed the realization of the latent processes in (1) and (2), that is if an individual in each province participates or not, and, for those that participate, if they have an employment or are unemployed:

[2] Notice that given the focus of this study, we deliberately exclude all determinants of the individual probability of being unemployed (such as those considered in studies using aggregate regional data) other than those operating at the micro level. We assume that regional differences in the macro determinants will affect the impact of the individual characteristics in each region.

$$P_i^p = 1\left(\mathbf{Z}_i^p \boldsymbol{\gamma}^p + \nu_i^p > 0\right) \tag{3}$$

$$U_i^p = 1\left(\mathbf{X}_i^p \boldsymbol{\beta}^p + \varepsilon_i^p > 0\right)$$

Assuming $\boldsymbol{\nu}$~N(0,1) and $\boldsymbol{\varepsilon}$~N(0,1), the so-called Heckprobit model can be settle down as:

$$\text{prob}\left(P_i^p = 1\right) = \Phi\left(\mathbf{Z}_i^p \boldsymbol{\gamma}^p\right) \tag{4}$$

$$\text{prob}\left(U_i^p = 1\right) = \Phi\left(\mathbf{X}_i^p \boldsymbol{\beta}^p\right)$$

$$\text{corr}(\boldsymbol{\varepsilon}\, \boldsymbol{\nu}) = \rho^p$$

where $\boldsymbol{\Phi}$ is the cumulative normal distribution function. When $\rho \neq 0$, it can be said that unobservable characteristics affect both the probability of participating and the probability of unemployment, causing the estimates from the (standard) probit model for *U* to be inconsistent. In contrast, the heckprobit maximum likelihood estimation procedure exploits the information in (4) to provide consistent and asymptotically efficient estimates of the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ coefficients.[3]

The matrix of characteristics **Z** may contain any or all variables in **X***,* though to prevent the identification for estimating the parameters based solely on the nonlinearity of the functional form, it is required that at least one of the variables included in **Z** is excluded from **X**, which means that it is assumed not to exert a direct effect on the probability of unemployment.

To assess the impact of individual's schooling in each province, we use the estimates of the above-mentioned coefficients to compute the corresponding marginal effect. More precisely, we compute the change caused by an additional year of schooling on the probability of unemployment conditional to participate in the labour market. Such conditional probability being defined as:

$$\text{prob}\left(U_i^p = 1 \mid P_i^p = 1\right) = \frac{\text{prob}\left(U_i^{\square} = 1\,,\, P_i^p = 1\right)}{\text{prob}\left(P_i^p = 1\right)} = \frac{\Phi_b\left(\mathbf{X}_i^p \boldsymbol{\beta}^p, \mathbf{Z}_i^p \boldsymbol{\gamma}^p, \rho^p\right)}{\Phi\left(\mathbf{Z}_i^p \boldsymbol{\gamma}^p\right)} \tag{5}$$

[3] We used the `heckprob` command in Stata 12 to obtain the estimates in the following section.

where $\mathrm{prob}(U_i^p = 1 \,,\, P_i^p = 1) = \Phi_b(\mathbf{X}_i^p \boldsymbol{\beta}^p, \mathbf{Z}_i^p \boldsymbol{\gamma}^p, \rho^p)$ is the joint probability of being unemployed and active, and $\boldsymbol{\Phi}_b$ denotes the cumulative distribution function of the bivariate normal. The marginal effect of education in each province is thus computed by averaging the estimated marginal effect of individuals in the sample of the corresponding province. The use of the particular estimated coefficients in (4) and the characteristics of individuals in each province prevent imposing the same impact of education regardless of the spatial location of the individual. On the contrary, this approach allows analysing the effect of education on unemployment from a spatial perspective.

As a final step, the empirical model in (4) is used to construct counterfactual unemployment rates in each province under different scenarios regarding the level of educational attainment. The sample average of the fitted conditional probabilities of unemployment in a province coincides with the ratio of unemployed individuals over the number of active individuals in that province, and thus with its actual unemployment rate. Counterfactual unemployment rates can thus be obtained by assigning to individuals in a province a given number of years of schooling instead of the actual ones in (5). For example, defining $\mathbf{X}_{C\_schSp}^p$ and $\mathbf{Z}_{C\_schSp}^p$ as the matrices $\mathbf{X}^p$ and $\mathbf{Z}^p$ in which the actual years of schooling for each individual in province p have been substituted by the country's average years of schooling, the resulting counterfactual unemployment rate in province p ($Ur_{C\_schSp}^p$) is obtained as:

$$Ur_{C_{sc}hSp}^p = \mathrm{prob}(U_i^p = 1 \mid P_i^p = 1)_{C_{sc}hSp} = \frac{\Phi_b(\mathbf{X}_{C_{sc}hSp}^p \hat{\boldsymbol{\beta}}^p, \mathbf{Z}_{C_{sc}hSp}^p \hat{\boldsymbol{\gamma}}^p, \hat{\rho}^p)}{\Phi\left(\mathbf{Z}_{C_{sc}hSp}^p \hat{\boldsymbol{\gamma}}^p\right)} \qquad (6)$$

where the ^ over the coefficients denotes the ML estimates.

Comparison of the distribution of actual unemployment rates with counterfactual distributions will allow us to assess the impact of the spatial sorting of individuals' education on inequality and spatial dependence of provincial unemployment rates.

6. RESULTS

6.1. Estimation of the impact of education on the probability of unemployment in Spanish provinces.

As indicated in the previous section, the ML estimates of the coefficients in (4) are used to compute the conditional marginal effects of the variables in **X** and **Z** for each province and time period under analysis. In other words, we have estimated the β, γ, and ρ coefficients for each of the 50 provinces in each of the two periods. These estimates were used then to compute the marginal effects in each province and period. In addition to the years of schooling, the factors included in **X** were gender, a set of age dummies, civil status (if the individual is married or not), a dummy distinguishing between natives and immigrants, and a variable accounting for the year (since as indicated in section 3 we pooled observations for two years in each period under analysis). All these factors were also included in **Z**, though to improve identification we also included other variables that are supposed to affect individual's decision to participate in the labour market, such as the number of household members, the number of children under 10 years old, the number of other members with an employment, and a dummy that indicates if the individual is the head of the household.

In order to save space we do not reproduce here the details of the one hundred estimates (50 provinces in the two periods), but just reproduce and discuss in some detail the conditional marginal effect associated to years of schooling.[4] In any case, it is worth mentioning that the standard statistics confirm that the coefficients of the set of factors included in **X** and **Z** contribute to explain the probability of unemployment in all provinces for both periods, and that the likelihood ratio rejects the null hypothesis of ρ=0 in most provinces, and thus that the sample selection probit specification in (4) is more appropriate than the simple probit model that does not control for participation in the labour market.[5]

---

[4] The full set of results are available from the authors upon request.

[5] It could be argued that the marginal effect from the simple probit model should be reported and use in the subsequent analyses for those provinces in which the LR test does not reject the null of ρ=0. Still, in those cases we have preferred to report and use the ones based on estimation of the sample selection probit in the sake of homogeneity, and because estimates from this model remain consistent regardless of the significance of ρ. Actually, the comparison of the marginal effect of education from the two estimates reveals that they are almost identical in those cases.

The conditional marginal effects of schooling in each province and time period are shown in Table A2 of the Appendix. To ease the interpretation of results, we have estimated the density of the distribution of these marginal effects for both periods. Results are displayed in Figure 5. It can be observed that the impact of education on the probability of unemployment is negative in both periods, although the magnitude is much higher in the one of high unemployment, 2011-12. Actually, the distribution of the marginal effect in the first period, 2006-07 is mostly concentrated in the interval -0.01 to 0. That is to say, for most provinces an additional year of schooling of the active population decreased the chances to be unemployed by a maximum of 1 percentage point (pp), and the mode is around 0.5pp. Interestingly, there are a few number of provinces (corresponding to the mass of probability for values of the marginal effect below -0.01) in which the impact is somewhat stronger than in the rest. The comparison of the values for 2006-07 in Tables A1 and A2 reveals that those provinces with the highest impact of education are the ones with the highest unemployment rates in that period (Badajoz, Cádiz, Córdoba, and Jaen, provinces in the South West of Spain).

In sharp contrast, the distribution of marginal effects in 2011-12 is far more disperse, corresponding to a moderate impact in some provinces and a much intense one in some others, with the mode slightly below -0.02. In any case, the distribution of the estimated marginal effects for 2011-12 is placed at the left of the distribution for 2006-07, which indicates that the effect of individuals' schooling on the chances to be unemployed is much higher in the period of crisis with unemployment rates far above those in 2006-07.

The association between the impact of schooling and unemployment rates is also derived from the spatial distribution of both magnitudes. Comparing the maps in Figure 6 for educational attainment in each province to those in Figure 2 for provincial unemployment rates, reveals that with few exceptions, the impact of schooling is higher in provinces experiencing high unemployment rates; the opposite being also true. Actually, this piece of evidence suggests that the return, in terms of chances to be unemployed, an individual can get from its investment in education is higher in areas of high unemployment. Or put it in other terms, low educated individuals face higher chances of unemployment, with respect to individuals with higher levels of educational attainment, in provinces characterised by high unemployment figures. In addition, figures on years of schooling in each province described in section 3 (last two columns of Table A1 and maps in Figure 4) points to a higher impact of education in provinces in which this type of capital is in short supply, and to a lower impact in places where educational attainment is higher (in agreement with a sort of decreasing return mechanism).

Finally, the maps in Figure 6 indicate that the distribution of the marginal effect of education is spatially clustered. This feature is confirmed by the results of the I-Moran test in the bottom panel of Table 1. It is observed that the null hypothesis of absence of spatial dependence is clearly rejected for both periods using the contiguity- and the distance-based weight matrices.

Summing up, the estimates of the conditional marginal effect of schooling in each province confirm that Spanish provinces do not just differ in the level of educational attainment of their active population (spatial sorting of schooling), but that they do also differ in the impact education has on the probability to be unemployed. Besides, the spatial analysis indicates that there is a close relationship between unemployment rates, level of educational attainment, and the impact of education in the Spanish provinces. Among other interesting issues, results confirm that improvements in the educational attainment of the active population would be an effective policy in provinces with high unemployment rates and low levels of education, particularly in periods of recession.

6.2. Counterfactual provincial unemployment rates.

Applying the strategy outlined at the end of section 5 we have obtained counterfactual unemployment rates for the Spanish provinces under two different scenarios. A first one in which we imposed that the years of schooling of individuals participating in the labour market in every province equal the average number of years of schooling in the country as a whole. The average years of schooling in the sample of the active population in Spain was 10.39 in 2006-07, and 10.75 in 2011-12. That is to say, broadly speaking the Spanish average was about 1 year higher than the average in provinces with the lowest endowment of education, and 1 year below the average in provinces with the highest endowment (see the last two columns of Table A1). The second scenario is a more dramatic one, since we imposed the average schooling observed for the province with the highest endowment, which was Guipúzcoa in the two periods under analysis with 11.69 and 12.19 years respectively. The gap between this province and the ones with the lowest levels of education was around 2.5 years, that with no doubt represents a substantial difference in the level of educational attainment of the active population.[6]

The distribution of the two counterfactual unemployment rates in each period is summarised by the estimated density functions in Figure 7. The dashed line represents the one for the counterfactual obtained when imposing the average years in Spain, while that obtained when using the average in Guipúzcoa (province with highest schooling) is represented by the dotted line. To ease the comparison we have also plotted, with a continuous line, the densities for the actual unemployment rates discussed in section 4. It is clearly observed that homogenizing the educational attainment of the active population causes a shift to the left in the unemployment distribution that, as expected, it is more dramatic the larger the increase in the years of schooling (counterfactual using the highest level of schooling).

As for the situation in 2006-07, the mass of probability in the upper part of the distribution (the highest unemployment rates) for the actual distribution vanishes in the counterfactual distributions. On the other hand, the mass of probability corresponding to low unemployment rates increases in the counterfactual distributions. It is obvious as well that the shift to the left is larger when homogenising using the highest levels of schooling. However, the degree of dispersion in the counterfactual distributions is similar to that for actual unemployment rates, even when conditioning using the level of education observed in the province with the highest endowment. Therefore, spatial sorting of education seems not to be the major cause of the dispersion observed in the actual provincial distribution of unemployment rates in that period.

In contrast, the comparison of the actual and counterfactuals distributions in 2011-12 suggests that a big deal of disparities across provinces in unemployment rates can be explained by the spatial sorting of education. It is clearly observed that the mass of probability associated to the very high levels of provincial unemployment rates

[6] To facilitate interpretation is useful to say that the magnitude of this gap is similar to that observed in the average years of schooling between Spain and Norway, which is the country with the highest value worldwide.

disappears when homogenising the level of education across provinces. Actually, the clear bimodality observed in the actual distribution does not show up in the counterfactuals, particularly when imposing the maximum value observed for the provincial average years of schooling. All in all, it can be said that the counterfactual distributions are far more concentrated than the actual distribution, and that they are located in a range of lower unemployment rates. In addition, results for the two periods under analysis lead us to conclude that the role played by the spatial sorting of education, i.e. by disparities across provinces in the level of educational attainment of the active population, is far more important in periods of high unemployment when the spatial gap (in absolute terms) in unemployment rates increases.

Finally, the representation of the counterfactual unemployment rates in a map reveals an interesting feature, which is the persistence of the spatial divide of Spain in terms of the incidence of unemployment even after homogenising the level of education of the active population. The corresponding maps for the two periods in the case of using the highest observed average schooling are reproduced in Figure 8.[7] In these maps we have kept the same range of values for each of the four categories (those corresponding to the quantiles in the actual distribution of unemployment rates in Figure 2) to facilitate the comparison between the spatial distribution of the actual and counterfactual unemployment rates. Especially in the period of high unemployment, 2011-12, the North-South divide is even more evident after netting out the effect of individuals' education in each province. This result thus suggests that spatial sorting of education accounts for an important amount of disparities across provinces, though it does not seem to be the responsible of the pattern observed in the spatial distribution of unemployment rates. Actually, the I-Moran tests for the counterfactual rates in the last set of rows in Table 1 indicate that the intensity of spatial dependence is similar to that observed for actual unemployment rates.

[7] The ones using the Spanish average years of schooling are not supplied to save space but are available upon request. In any case, the main picture derived from these results is robust to the use of any of the measures.

7. CONCLUSIONS

The existing literature on the determinants of regional unemployment obtained using aggregate data is inconclusive on the impact of education on unemployment rates. López-Bazo and Motellón (2012) argued that the lack of robustness in the results might be caused, precisely, by the use of aggregate information, and suggested exploiting micro-data to obtain founded conclusions on this issue. In fact, they showed that an important part of the gap observed between the groups of regions with the highest and lowest unemployment rates in Spain can be explained by differences between the two groups in worker's characteristics, including the educational attainment of the individual.

In this paper we have derived complementary evidence on this issue. After showing that the provincial distribution is rather disperse and spatially correlated, particularly in periods of high unemployment, our results have revealed that provinces also differ markedly in the level of educational attainment of the active population, and in the impact education has on the individual's probability of being unemployed. They also point to a much higher effect of education during a recession, in which unemployment rates increase substantially.

Results from the counterfactual "*what if*" exercise using the particular estimate of the impact of schooling in each province allows us to conclude that unemployment rates in Spanish provinces will be not only lower but also less scattered, if the level of education is more homogeneous across provinces. This supports the implementation of policies aiming at stimulating education in provinces historically suffering high unemployment, and also in those in which the labour demand by the construction sector was particularly high during the expansion period (such as those in the Mediterranean coast). Such a high demand caused that a large percentage of the youth population left the education system at early stages to occupy unskilled jobs in activities related to the building sector. The collapse of that sector after the crisis expelled those low educated workers, leaving them with scarce expectations of employability.

REFERENCES

Bande R, Fernández M, Montuenga V (2008) Regional unemployment in Spain: Disparities, business cycle and wage setting. *Labour Economics* 15: 885-914

Bande R, Karanassou M (2009) Labour market flexibility and regional unemployment rate dynamics: Spain 1980–1995. *Papers in Regional Science* 88: 181-207

Blanchard O J, Katz L F (1992) Regional evolutions. *Brookings Papers on Economic Activity. Economic Studies Program. The Brookings Institution* 23: 1-76

Cracolici M F, Cuffaro M, Nijkamp P (2007) Geographical distribution of unemployment. An analysis of provincial differences in Italy. *Growth and Change* 38: 649-670

Decressin J, Fatás A (1995) Regional labor market dynamics in Europe. *European Economic Review* 39: 1627-1655

Elhorst J P (1995) Unemployment disparities between regions in the European Union. In: Armstrong H W, Vickerman R W (eds) *Convergence and divergence among European Regions*. Pion Ltd, London

Elhorst J P (2003) The mystery of regional unemployment differentials: Theoretical and empirical explanations. *Journal of Economic Surveys* 17: 709–748

Evans P, McCormick B (1994) The new pattern of regional unemployment: causes and policy significance. *Economic Journal* 104: 633–647

Filiztekin A (2009) Regional unemployment in Turkey. *Papers in Regional Science* 88: 863-878

Hall R E (1972) Turnover in the labor force. *Brookings Papers on Economic Activity* 3: 709–764

Lazar F (1997) Regional unemployment rate disparities in Canada: some possible explanations. *Canadian Journal of Economics* 10: 112–129

López-Bazo E, del Barrio T, Artís M. (2002) The regional distribution of Spanish unemployment: A spatial analysis. *Papers in Regional Science* 81: 365-389

López-Bazo E, del Barrio T, Artís M (2005) Geographical distribution of unemployment in Spain. *Regional Studies* 39: 305-318

López-Bazo E, Motellón E (2012) The regional distribution of unemployment. What do micro-data tell us? *Papers in Regional Science*, Early view, doi: 10.1111/j.1435-5957.2012.00456.x

Marston S T (1985) Two views of the geographic distribution of unemployment. *Quarterly Journal of Economics* 100: 57-79

Martin R (1997) Regional unemployment disparities and their dynamics. *Regional Studies* 31: 237-252

Martin R, Sunley P (1999) Unemployment flow regimes and regional unemployment disparities. *Environmental and Planning A* 31: 523–550

Mincer J (1991) Education and unemployment. *NBER Working Paper* No. W3838

OECD (1989) Regional unemployment in OECD countries. *Employment Outlook.* Paris

OECD (1990) Supply and demand in regional labour markets: population growth, migration, participation, and earnings differentials. *Employment Outlook*. Paris

OECD (2000) Disparities in regional labour markets. *Employment Outlook*. Paris

OECD (2005) How persistent are regional disparities in employment? The role of geographic mobility. *Employment Outlook*. Paris

OECD (2011a) Regions at a glance 2011. Paris.

OECD (2011b) *Education at a glance 2011: OECD indicators*. Paris.

Overman H G, Puga D (2002) Unemployment clusters across Europe's regions and countries. *Economic Policy* 34: 115-147

Partridge M D (2001) Exploring the Canadian-U.S. unemployment and nonemployment rate gaps: Are there lessons for both countries?. *Journal of Regional Science* 41: 701-734

Partridge M D, Rickman D S (1997) The dispersion in US unemployment rates: The role of market and nonmarket equilibrium factors. *Regional Studies* 31: 593–606

Patacchini E, Zenou Y (2007) Spatial dependence in local unemployment rates. *Journal of Economic Geography* 7: 169-191

Rosen S (1974) Hedonic prices and implicit prices. *Journal of Political Economy* 86: 34–53

Taylor J, Bradley S (1997) Unemployment in Europe: A comparative analysis of regional disparities in Germany, Italy and the UK. *Kyklos* 50: 221–245

Van de Ven, W P M M, Van Praag B M S (1981) The demand for deductibles in private health insurance : A probit model with sample selection, *Journal of Econometrics* 17: 229-252

Table 1. Results of the I-Moran spatial dependence test for the Spanish provinces.

	Contiguity	Inverse sq. distance
Unemployment rate		
2011-12	0.714***	0.583***
2006-07	0.733***	0.488***
Years of schooling		
2011-12	0.371***	0.436***
2006-07	0.316***	0.400***
Marg. Eff. of Schooling		
2011-12	0.659***	0.459***
2006-07	0.589***	0.362***
Counterfact. Unemp rate (highest schooling)		
2011-12	0.687***	0.546***
2006-07	0.691***	0.428***

Note: *** denotes p-value < 0.01.

Figure 1. Estimated density functions of unemployment rates in Spanish provinces, 2011-12 and 2006-07.

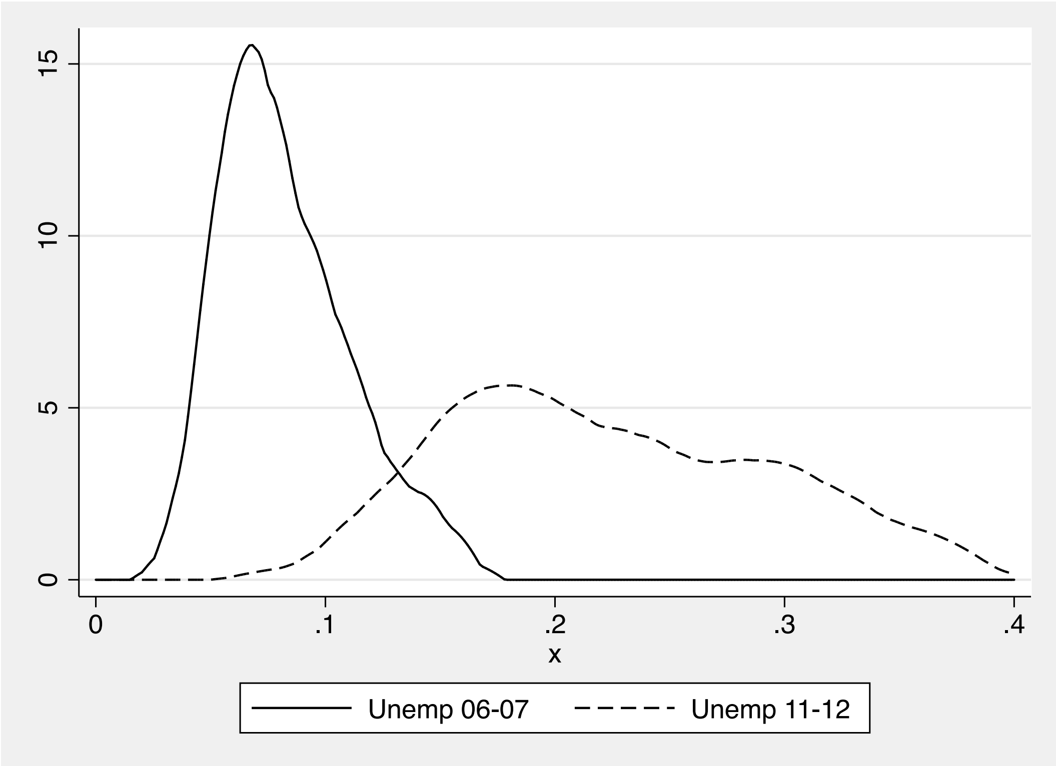


Figure 2. Estimated density function of years of schooling in Spanish provinces, 2011-12 and 2006-07.

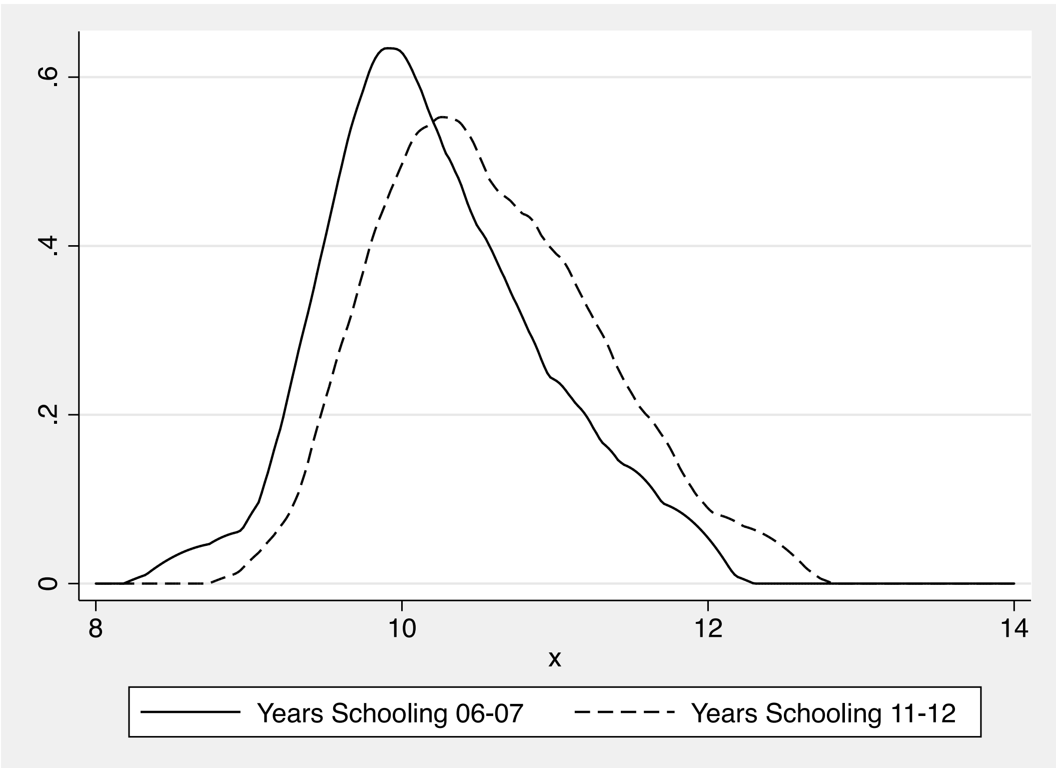
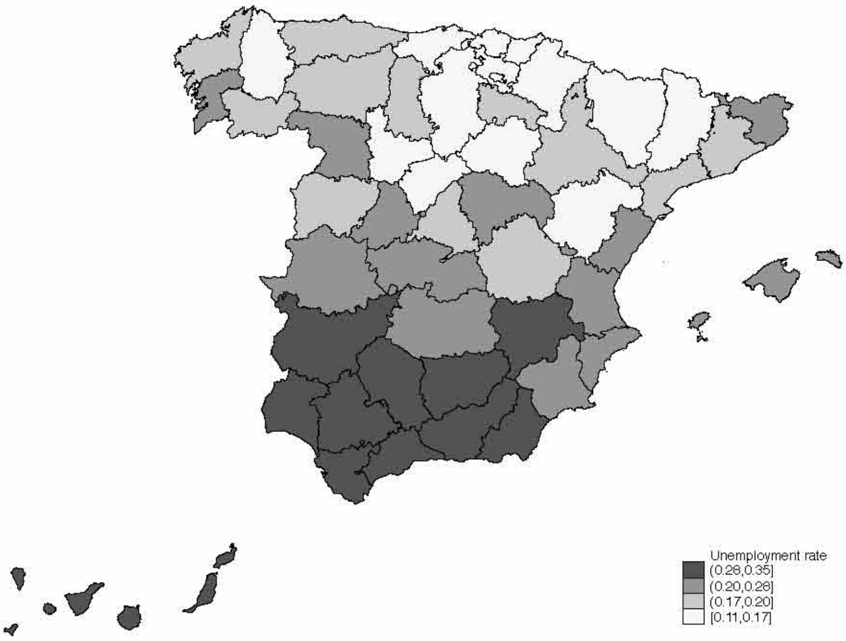


Figure 3. Unemployment rates in Spanish provinces.

2011-12



2006-07

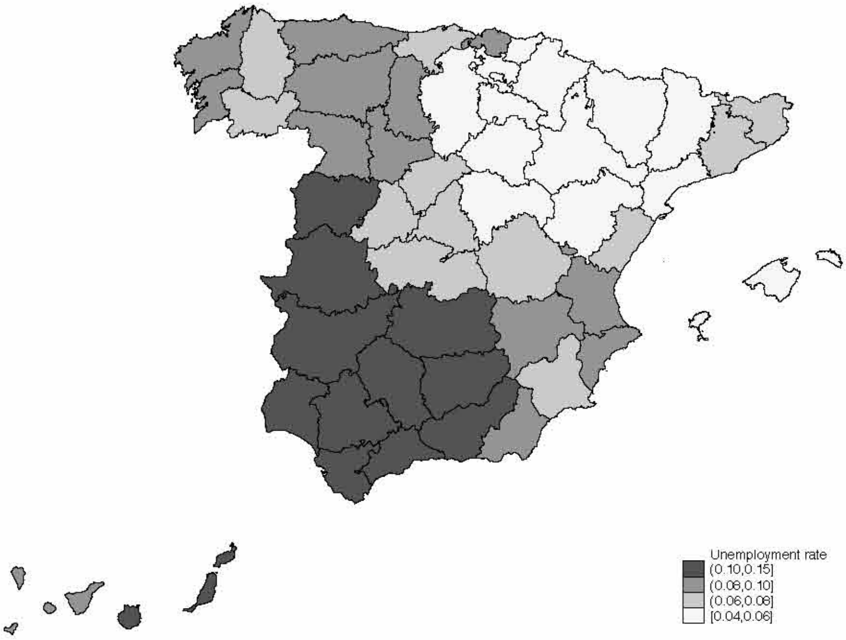
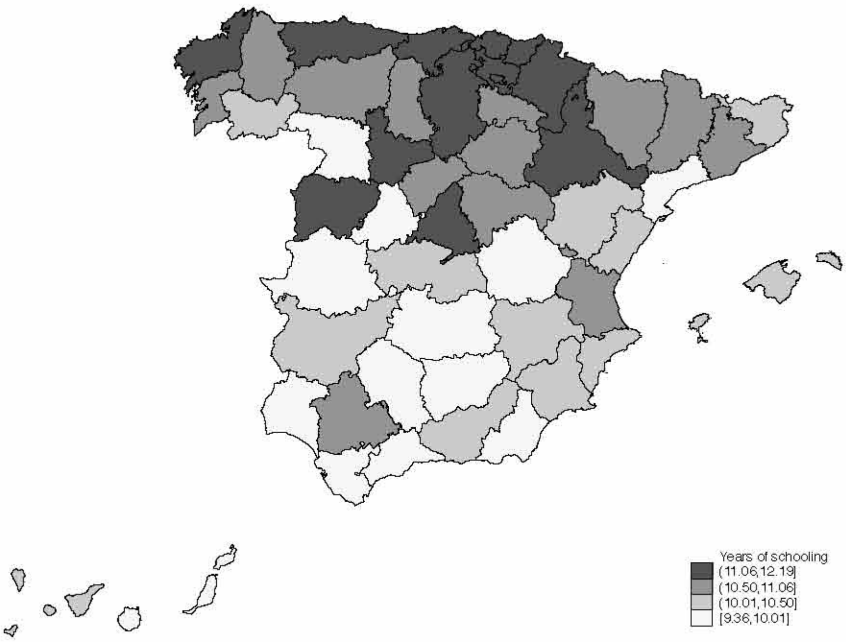


Figure 4. Years of schooling in Spanish provinces.

2011-12



2006-07

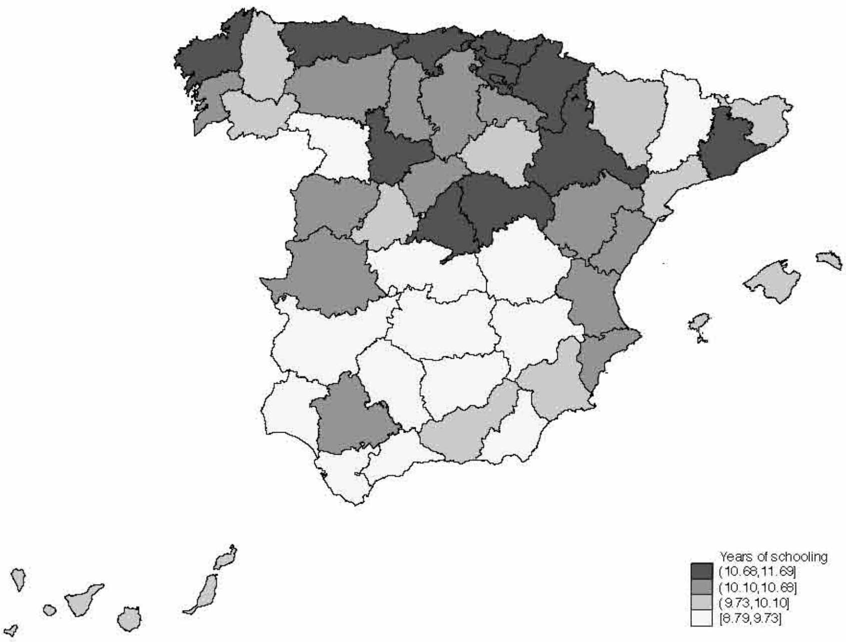


Figure 5. Estimated density function of marginal effect for years of schooling in Spanish provinces, 2011-12 and 2006-07.

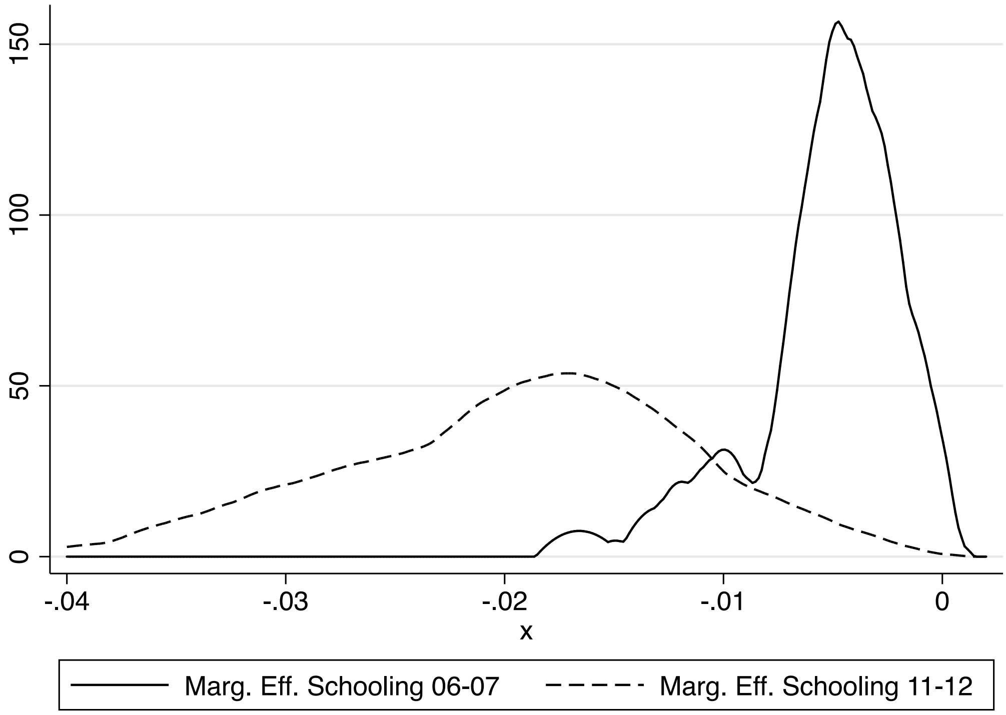
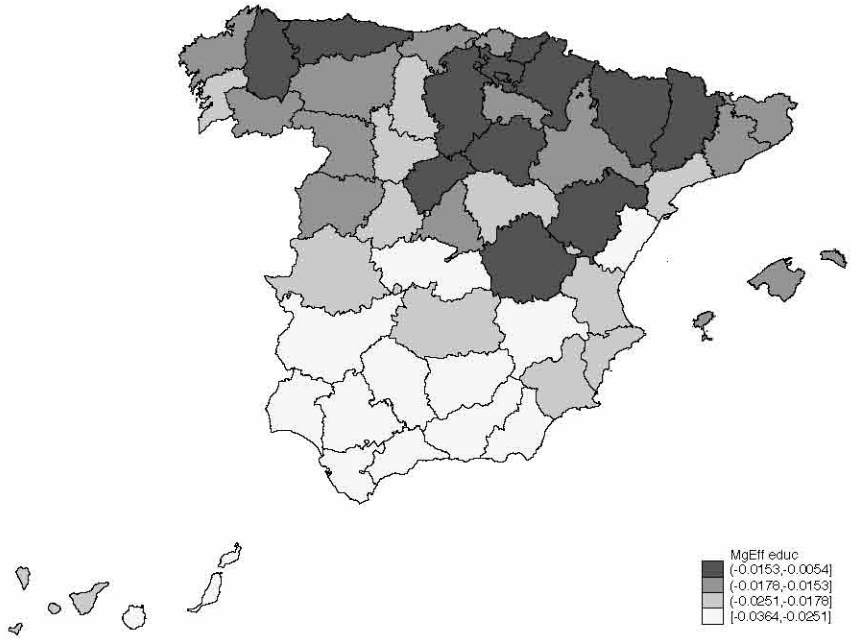


Figure 6. Marginal effect of schooling in Spanish provinces.

2011-12



2006-07

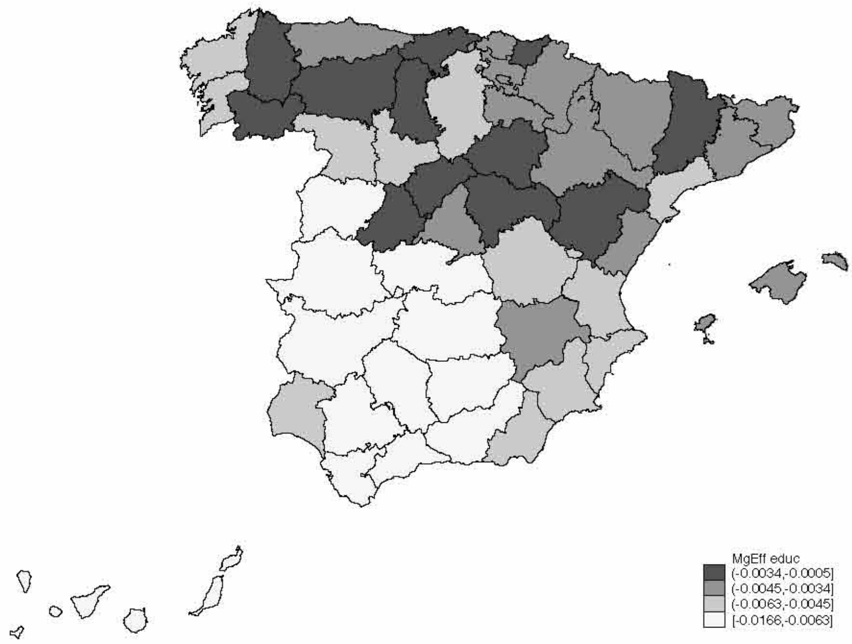
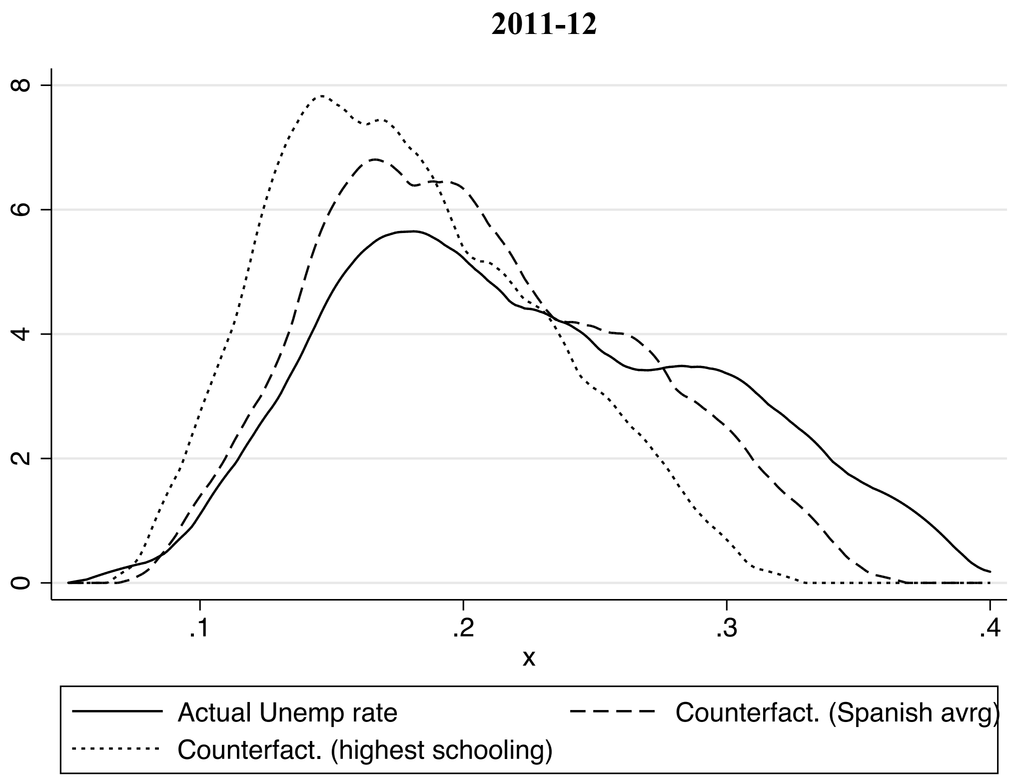


Figure 7. Estimated density function for actual and counterfactual unemployment rates in Spanish provinces.



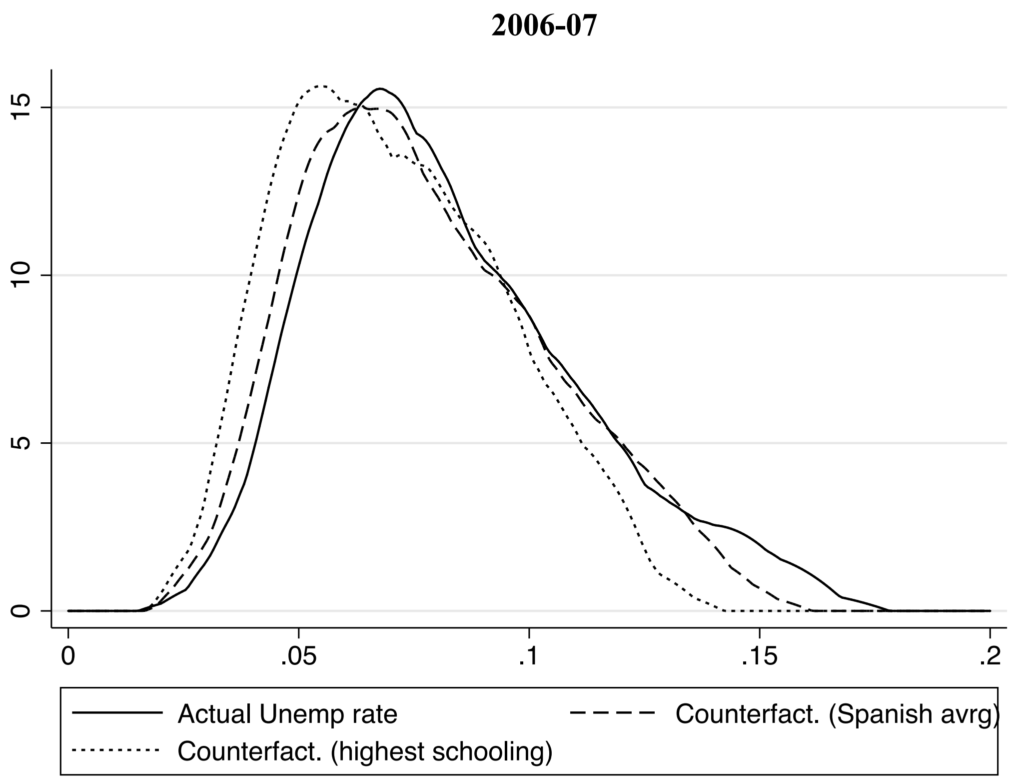
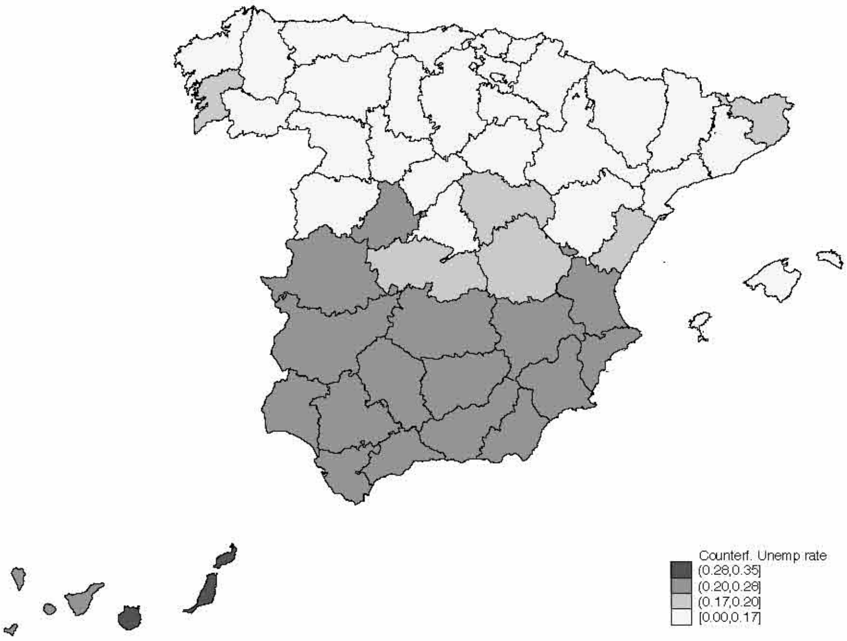


Figure 8. Counterfactual unemployment rates (highest schooling) in Spanish provinces.

2011-12



2006-07

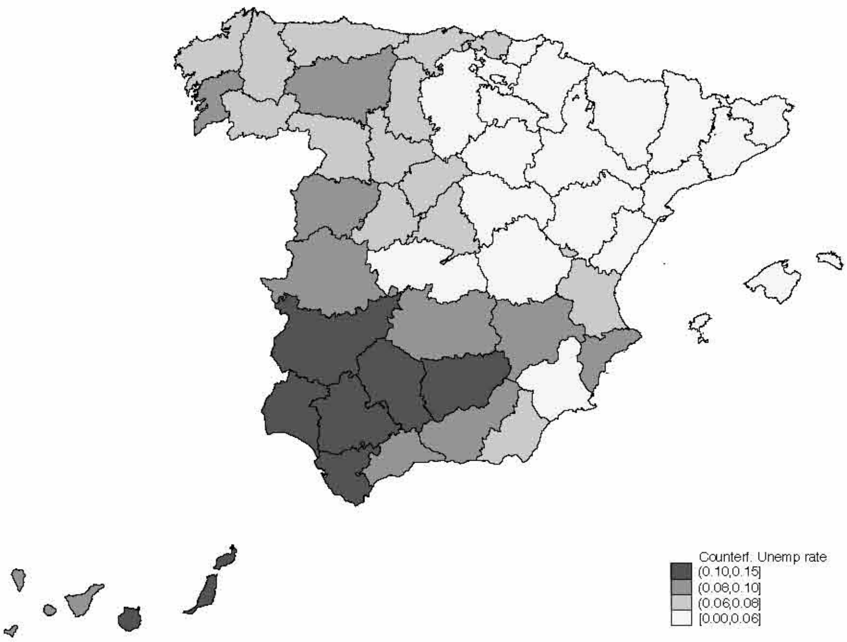


Table A1. Unemployment rate and years of schooling in Spanish provinces.

	Unemployment rate		Schooling	
	2011-12	**2006-07**	**2011-12**	**2006-07**
Álava	0.1183	0.0565	11.60	11.54
Albacete	0.2794	0.0994	10.38	9.73
Alicante	0.2546	0.0943	10.28	10.12
Almería	0.3547	0.1000	9.36	9.35
Ávila	0.2566	0.0636	9.86	9.84
Badajoz	0.2942	0.1422	10.15	9.66
Islas Baleares	0.2045	0.0610	10.32	9.91
Barcelona	0.1995	0.0640	10.92	10.81
Burgos	0.1588	0.0630	11.19	10.65
Cáceres	0.2628	0.1086	9.90	10.11
Cádiz	0.3320	0.1534	9.94	9.68
Castellón	0.2782	0.0676	10.29	10.21
Ciudad Real	0.2789	0.1005	9.95	9.67
Córdoba	0.3363	0.1430	10.01	9.56
La Coruña	0.1751	0.0831	11.29	10.84
Cuenca	0.1988	0.0692	9.54	8.79
Gerona	0.2386	0.0634	10.33	9.83
Granada	0.3307	0.1110	10.49	10.00
Guadalajara	0.2185	0.0607	10.65	10.71
Guipúzcoa	0.1110	0.0503	12.19	11.69
Huelva	0.2902	0.1152	9.96	9.58
Huesca	0.1494	0.0450	10.80	10.03
Jaén	0.3344	0.1298	9.79	9.67
León	0.1907	0.0842	10.83	10.35
Lleida	0.1587	0.0507	10.57	9.54
La Rioja	0.1972	0.0556	11.06	10.54
Lugo	0.1434	0.0651	10.51	9.74
Madrid	0.1755	0.0666	12.10	11.58
Málaga	0.3253	0.1058	9.88	9.68
Murcia	0.2541	0.0724	10.01	9.81
Navarra	0.1474	0.0545	11.40	11.11
Orense	0.1900	0.0734	10.28	9.96
Asturias	0.1920	0.0863	11.31	10.70
Palencia	0.1885	0.0793	10.63	10.17
Las Palmas	0.3479	0.1121	9.70	9.88
Pontevedra	0.2256	0.0871	10.54	10.22
Salamanca	0.1814	0.1008	11.37	10.68
Santa Cruz de Tenerife	0.2850	0.1002	10.38	10.04
Cantabria	0.1622	0.0656	11.24	10.81
Segovia	0.1579	0.0696	10.97	10.34
Sevilla	0.2845	0.1267	10.59	10.37
Soria	0.1407	0.0506	10.76	10.09
Tarragona	0.2007	0.0630	9.99	9.81
Teruel	0.1562	0.0400	10.16	10.30
Toledo	0.2659	0.0751	10.10	9.56
Valencia	0.2498	0.0790	10.90	10.52
Valladolid	0.1745	0.0798	11.42	11.08
Vizcaya	0.1481	0.0786	12.02	11.57
Zamora	0.2052	0.0898	9.57	8.94
Zaragoza	0.1922	0.0599	11.19	11.09

Table A2. Marginal effects for years of schooling in the Spanish provinces from the heckprobit model.

	2011-12	2006-07
Álava	-0.0107	-0.0042
Albacete	-0.0317	-0.0036
Alicante	-0.0202	-0.0058
Almería	-0.0256	-0.0063
Ávila	-0.0218	-0.0015
Badajoz	-0.0273	-0.0166
Islas Baleares	-0.0167	-0.0038
Barcelona	-0.0166	-0.0042
Burgos	-0.0148	-0.0059
Cáceres	-0.0230	-0.0107
Cádiz	-0.0257	-0.0125
Castellón	-0.0268	-0.0044
Ciudad Real	-0.0238	-0.0063
Córdoba	-0.0364	-0.0133
La Coruña	-0.0164	-0.0045
Cuenca	-0.0073	-0.0051
Gerona	-0.0166	-0.0034
Granada	-0.0308	-0.0089
Guadalajara	-0.0188	-0.0033
Guipúzcoa	-0.0068	-0.0029
Huelva	-0.0314	-0.0063
Huesca	-0.0128	-0.0036
Jaén	-0.0356	-0.0111
León	-0.0156	-0.0010
Lleida	-0.0114	-0.0025
La Rioja	-0.0164	-0.0036
Lugo	-0.0054	-0.0021
Madrid	-0.0166	-0.0040
Málaga	-0.0271	-0.0095
Murcia	-0.0208	-0.0058
Navarra	-0.0085	-0.0041
Orense	-0.0169	-0.0025
Asturias	-0.0138	-0.0036
Palencia	-0.0196	-0.0012
Las Palmas	-0.0251	-0.0063
Pontevedra	-0.0184	-0.0048
Salamanca	-0.0169	-0.0065
Santa Cruz de Tenerife	-0.0226	-0.0070
Cantabria	-0.0159	-0.0015
Segovia	-0.0114	-0.0005
Sevilla	-0.0313	-0.0097
Soria	-0.0091	-0.0013
Tarragona	-0.0195	-0.0055
Teruel	-0.0134	-0.0010
Toledo	-0.0288	-0.0066
Valencia	-0.0214	-0.0053
Valladolid	-0.0191	-0.0058
Vizcaya	-0.0153	-0.0035
Zamora	-0.0158	-0.0052
Zaragoza	-0.0172	-0.0035