40th ERSA Meeting

THE GEOGRAPHICAL DISTRIBUTION OF UNEMPLOYMENT IN SPAIN

Enrique López-Bazo
Tomás del Barrio
Manuel Artís

Research Group "Anàlisi Quantitativa Regional"
Dpt of Econometrics, Statistics and Spanish Economics
University of Barcelona
Address: Avda Diagonal 690, ES-08034 Barcelona, Spain
Phone: +34 93 4024320
Fax: +34 93 4021821
e-mail: elopez@eco.ub.es barrio@eco.ub.es artis@eco.ub.es
URL: http://www.eco.ub.es/~elopez

ABSTRACT. An interesting feature of the dynamics of unemployment in the EU economies is that while the aggregate unemployment rates move with the business cycle, its regional distribution seems to remain quite stable or tends to a sort of polarisation. This paper discusses different hypotheses that have been proposed to explain heterogeneity in the spatial distribution of unemployment, and performs an empirical analysis for the Spanish provinces. The paper assesses the effects of skill and sectoral mismatch, and demographic factors in explaining spatial differences in unemployment rates. Although regional differences in sectoral specialisation and skills account for some of the dispersion in the distribution in the mid-eighties, they have no role in explaining the one today. In turn, demography became more important at the end of the period, though it is not able to explain completely the emergence of (at least) two clusters of regions with separate unemployment rates.

Keywords: unemployment rate differentials, regional labour markets, distribution dynamics, spatial analysis, Spain

JEL: E24, R12, R23

INTRODUCTION

High unemployment has been one of the main problems in most of the western economies in the last decades. This is particularly the case in some European countries such as Spain. The Spanish experience has focused the attention of the literature because of the extremely large unemployment rates and the persistence of shocks affecting the labour market. Structural conditions, rigidities and the system of unemployment benefits have been pointed as the factors that could explain why unemployment in Spain has shown such a peculiar behaviour when compared for instance with the one in Portugal (Blanchard and Jimeno, 1995).

Besides the nation-wide aggregate unemployment, another interesting but as far as we know less studied issue has been the geographical distribution of unemployment. There is some evidence on the relevance of spatial differentials in unemployment rates in the US, Canada and Europe. As will be shown later, the Spanish experience is again extreme. The distribution of unemployment rates is characterised by large differences and strong stability in the ranking of regions.

As a matter of example, the Spanish provinces (NUTS III regions in Spain[1]) with higher rates in 1999 were two regions in the South of the country: Cádiz (32.5%) and Córdoba (30.6%). They doubled the Spanish aggregate that year (15.8%). They were among the regions with higher rates in 1990 (32.9% and 25.8%), 1980 (24.2% and 14.1%) and 1976 (10.1% and 9.8%). Curiously, Cádiz in all these years and Córdoba in some of them also doubled the Spanish rates (16.2%, 11.4% and 4.5% respectively). In contrast, Lleida in the North-East and Soria in the Centre had rates even below the average in the EU in 1999 (5.5% and 6.3%). And they have also been among the most favoured provinces in Spain in the last decades, with rates that were around the half of the Spanish average.

On the other hand, regional economies seem to have different fortunes in the recent trend towards lower levels of aggregate unemployment. The Spanish rate decreased from 24.1% in 1994 to 15.8% in 1999. While Alava (22.3% to 10.6%), Alicante (26.2% to 13.5%) and Barcelona (23.0% to 11.4%) saw more intense reductions, in some other provinces reductions were very limited (Cáceres 24.9% to 22.5%; Ciudad Real 20.0% to 17.3%). What is more striking, there are provinces in which rates stagnated or even increased over that period (Córdoba 30.5% to 30.6%; Ourense 14.6% to 20.4%). This evidence might point to some sort of polarisation by which some regions are approaching unemployment rates around the average in the EU, while some others lay aside this process.

[1] See the Appendix for further details.

Considering this simple evidence, our first aim is the analysis of the evolution of the geographical distribution of the unemployment rates in Spain. The shape of the distribution is supposed to provide evidence not only on the amount of the disparities, but also on the formation of clusters of regions with separate unemployment rates. Additionally, the study of the dynamics within the distribution will provide straight evidence on the persistence of regional differentials as well as some intuition on the more likely evolution in the years to come. In a second step, the contribution of some factors --which can affect the level of unemployment on a priori grounds-- to the dynamics of the geographical distribution of unemployment in Spain, will be assessed. Given our interest in the whole distribution we cannot constrain the study to the traditional regression analysis. Instead, we will compare the entire observed distribution to the one obtained once the effects of some determinants of the unemployment have been removed.

We depart from some recent studies that have focused the attention in how fast labour markets variables adjust to shocks in the European economies (Decressin and Fatás, 1995; Jimeno and Bentolila, 1998). Here we will try to assess how much of the features observed in the spatial distribution of the unemployment rates can be explained by some factors that can potentially affect unemployment. To some extent, results for the Spanish economy can shed light on the processes in course in some other European economies and in the entire EU. In this sense, Overman and Puga (1999) have provided results on similar issues for a set of EU regions, thought with some methodological differences.

Why do regions differ so markedly in their unemployment rates? Or, in other words, why is the geographical distribution of unemployment so uneven? It is important to note that this question is not only important per se, but also because it might influence relative income per capita. That is to say, differentials in unemployment rates, jointly with differences in labour productivity and participation rates, translate into inequality in income per capita. In this sense, Esteban (1999) concludes that inequality in the entire EU is basically due to differences in productivity, with unemployment differentials having a low, although increasing, role. However, for some countries such as Italy and Spain he reports that regional differences in unemployment rates account for more that the 20% of the disparities in income per capita.

The rest of the paper is organised as follows. The next section briefly discusses some hypotheses that have been proposed to explain spatial differences in unemployment rates, and presents some related empirical evidence. Section 3 analyses the distribution of regional unemployment differentials in Spain over the period from 1985 to 1997. Once the distribution and its dynamics have been characterised, the role of some factors in explaining unemployment differentials is assessed in section 4, while results on the effect of unemployment rates in the neighbouring regions are summarised and discussed in section 5. Finally, section 6 concludes.

THE SPATIAL DISTRIBUTION OF UNEMPLOYMENT

Besides the traditional interest in the analysis of the evolution of the rate of unemployment, and the causes that could explain its anomalous high levels in some economies and periods[2], increasing attention has been devoted to the study of its geographical distribution. In a world characterised by the absence of adjustment costs and rigidities, one should expect no persistence in the difference between unemployment rates across locations. Excess labour in an area will disappear quickly due to workers' movements to lower unemployment areas. However, the evidence for several economies seems to indicate quite the opposite: high unemployment regions in a given decade do continue showing this qualification in the following decades, while the same rules for those with low unemployment.

Slow wage adjustments and large costs of migration are likely to explain why idiosyncratic shocks, or different regional responses to common shocks, might cause unemployment rates to differ markedly across regions for long. Under this explanation, heterogeneity in the spatial distribution of unemployment is a disequilibrium phenomenon as defined in Marston (1985). Three forces could drive a regional labour market towards the equilibrium: 1) out-migration because of the high unemployment rate, 2) fall in wages due to excess of labour supply, and 3) firms in-migration attracted by large unemployed labour force and low wages. As Marston points out, the important question is how strong these forces are. His results showed how they are strong enough in the US to restore spatial equilibrium within a year.

A second explanation for why geographic areas differ in their unemployment rates is suggested in Marston (1985) based on ideas in Hall (1972) and Rosen (1974). There is a steady-state relationship in unemployment rates across regions that is a function of their endowment of certain factors. Because the endowment differs among regions, the spatial distribution of unemployment is not homogeneous. Besides, as far as the endowment remains stable over time, the distribution of unemployment is not expected to change dramatically.

[2] Useful surveys include those by Bean (1994), Røed (1997).

This equilibrium hypothesis is then based on the idea that workers have incentives not to migrate when unemployed because they value those endowments in some sense. Evidence on high wages in high unemployment areas supports this viewpoint, as well as the preference for facilities and amenities. Martin (1997) extends the list of factors that could explain unemployment equilibrium differentials to permanent differences in economic, institutional and labour market characteristics across regions.

It is sensible to think that both phenomena simultaneously contribute to explain disparities in unemployment rates across economies. Thus, differences in the industrial mix, in matching efficiency of local labour markets, in several types of mismatch, and in the spatial distribution of amenities might, among other factors, induce the geographic distribution of unemployment in equilibrium. Likewise, forces pushing labour markets towards this equilibrium, when there is a temporary shock, might be weak enough not to avoid the system to be out of its long-run position for some time. In this case, it is likely that the observed differentials in unemployment rates across areas in any period were due to both factors. Empirical analyses may then face problems in isolating their contributions. However, to what extent each one of them is explaining the dispersion in the spatial distribution of unemployment has obvious implications when designing policies to alleviate such situation (Marston, 1985; Martin, 1997).

In its empirical study, Marston observed that shocks in the US metropolitan areas are eliminated within a year by mobility of workers across areas. Blanchard and Katz (1992) also point to the role of migration in returning the system of economies quickly to the relative unemployment equilibrium rates in the US States. In these studies, the steady-state relationship among regional unemployment rates is supposed to be some function of the amenities and land endowments in each area.

Although the mechanism of adjustment seems to come primarily by changes in participation, Decressin and Fatás (1995) conclude that the persistence of shocks to relative unemployment is even lower in Europe than in the US[3]. But this does not mean that differences in unemployment rates across the European regions do not show persistence. On the contrary, the authors conclude on the existence of high persistence of the regional unemployment differentials in Europe.

Additional evidence on the uneven spatial distribution of unemployment rates has been reported for Canada (Lazar, 1977) and some EU Member States (OECD, 1989; Evans and McCormick, 1994; Martin and Sunley, 1999). In all cases regional rankings seem to be quite stable. Overman and Puga (1999) report interesting evidence on some sort of polarisation in the unemployment regional distribution in the EU. If this were so, the only movements within the distribution would be to exacerbate differences between low and high unemployment regions that, besides, tend to be close together.

In this sense, it is well reported in the recent literature that the response of wages and migration to variations in unemployment is much higher in US than in Europe. As a consequence, it is more likely in the US that areas with large unemployment in a decade show low rates in the next one (Bertola and Ichino, 1996; Blanchard and Katz, 1992).

Undoubtedly, Spain is one of the countries in which unemployment has shown stronger persistence, and the regional ranking has experienced only very minor changes in the last decades. Not only real wages differentials are lower than disparities in unemployment rates, but their response to changes in regional unemployment is very low (Lorente, 1992; Jimeno and Bentolila, 1998). This is particularly the case when one controls for differences in the industrial mix and professional categories. The characteristics of the Spanish labour market and the process of centralised wage bargaining explains why local wages only marginally accommodates to changes in local unemployment.

Unlike the evidence for the EU regions, regional differences in participation rates seem to be very persistent. As a result neither migration nor adjustments in participation contribute to eliminate rapidly the effects of labour demand shocks (Jimeno and Bentolila, 1998).

Furthermore, migration flows, which acted as an important mechanism to balance the Spanish labour market up to the eighties, decrease to very low levels in the last two decades. The coincidence between the decrease in migration and the dramatic increase in the Spanish unemployment rate agrees with the idea of a negative relationship between both magnitudes (Pissarides and Wadsworth, 1989; Pissarides and McMaster, 1990). Recent empirical models of migration behaviour for the Spanish economy do support this hypothesis (i.e. Bentolila and Dolado, 1991)[4]. Besides, migration seems to be only marginally influenced by differences in unemployment rates across regions.

However, and despite the previous evidence pointing to the disequilibrium explanation for the Spanish economy, we cannot discard the simultaneous influence of equilibrium factors. Actually, Martin (1997) argues that persistence and stability of the regional

[3] However, large persistence is observed in Europe when they measure the response of absolute unemployment.
[4] However, as shown in de la Fuente (1999) for Spain, migration is actually sensitive to more appropriate measures of the probability of employment.

unemployment structure in UK is basically an equilibrium phenomenon. With this aim, in the next section we analyse the dynamics of the spatial distribution of unemployment in Spain in the last decades. Results confirm the strong stability and persistence of regional differentials. If something, there seems to be appearing a tendency towards polarisation in, at least, two groups: areas with large unemployment differentials and areas with unemployment rates some percentage points below the national average. Next, we assess the role of some factors in explaining the geographic distribution of unemployment. While in the mid-eighties sectoral composition and education of the labour force played a role, at the end of the nineties only demographic variables explain the dispersion in the distribution. Additionally, a larger proportion of the dispersion remains unexplained at the end of the period.

To end this section, we should say that our empirical analysis moves away from the traditional study of an (*average) representative economy*. As our main concern is to know what has happened with the geographic distribution of unemployment, inference at use on the *representative economy* is not of interest. Instead we will use tools that will allow us to say things on the shape of the distribution and the movements within it --over time or when conditioned to some of the factors that could explain unemployment differentials across the space. They have been proposed recently in different papers by Quah (1996, 1997) and applied in e.g. Bianchi (1997), Magrini (1999), and López-Bazo et al (1999) to study convergence in income per capita across countries and across regions. In this sense, as far as we know the approach applied here has only been used in Overman and Puga (1999) to study the distribution of unemployment rates in the EU regions.

DYNAMICS OF THE REGIONAL DISTRIBUTION OF UNEMPLOYMENT IN SPAIN[5]

The evolution of the aggregate Spanish unemployment has been profusely studied in the literature (Blanchard and Jimeno, 1995; Dolado and Jimeno, 1997; Marimon and Zilibotti, 1998). The average rate remains stable around 2-3% during the sixties. It increased moderately the next decade to around 10% at the beginning of the eighties (Table 1). Then, the unemployment rate doubled in five years to a situation in which more than 20 out of 100 workers were unemployed. Later, unemployment rates moved together with the business cycle, but always in a range far above the rates in other western economies. In this period, the

[5] All the results were obtained by using GAUSS v3.2.38. Data and codes are available from the authors upon request.

standard deviation, as a raw measure of unemployment differentials in the Spanish provinces (NUTS3)[6], increased notably up to the mid-eighties. Afterwards, it also shifted with the business cycle[7].

Table 1. Evolution of the unemployment rate in Spain.

	nation-wide aggregate	Stand dev of provinces	maximum	minimum
1978	6.97	3.83	15.55	1.62
1985	21.63	6.80	33.59	6.61
1990	16.21	6.60	32.92	3.22
1997	20.80	6.99	38.21	8.36

Notes: maximum and minimum correspond to the unemployment rates in the provinces with higher and lower rates each year. Then, they do not necesarily belong to the same province each year.

The comparison of unemployment rates in provinces with extreme values each year provides a clear picture of the magnitude of the spatial differences. If we use unemployment rates as a rough measure of the probability of being unemployed, figures in the last two columns of Table 1 do not deserve additional comments. Actually, the more recent figures seem to indicate that some provinces in the north-east of Spain are in a situation of full employment while in some hundreds of kilometres rates remain above 20%.[8]

We are going to focus our attention in the rest of the paper on the distribution for two particular time periods: 1985 and 1997. They cover approximately an entire cycle, and the average unemployment rates in these years were roughly similar. Therefore, in some sense, the effects of asymmetries in the regional response to phases in the cycle are minimised. Additionally, similar unemployment rates for the aggregate in both periods cause the analysis of relative or absolute deviations not to differ markedly[9].

Figure 1 plots de estimated density function of the distribution of unemployment rates in the Spanish provinces. It proxies the shape of the distribution, and in essence provides more information than the single measures of position or dispersion. What this figure allows

---

[6] A brief description of the units of analysis in this study is provided in the Appendix.

[7] As already reported for some other economies, absolute and relative unemployment differentials move in opposite directions, being the former (latter) positively (negatively) related to the aggregate unemployment rates. In our case, dispersion in relative unemployment decreased up to 1985, showing a countercyclical behaviour afterwards.

[8] The fact that unemployment differentials across Spanish regions persist despite the continuous decrease in the average rate has recently been noted by government agencies as one of the main drawbacks of the Spanish economy.

[9] See Martin (1997) for a discussion of regional unemployment disparities in terms of *relativities* or *differentials*. As a matter of comparison, Overman and Puga (1999) analyse the ratio of the regional unemployment rate to the EU wide average unemployment rate. We checked the sensitivity of our results to *relativities*. The main conclusions remained unaffected.

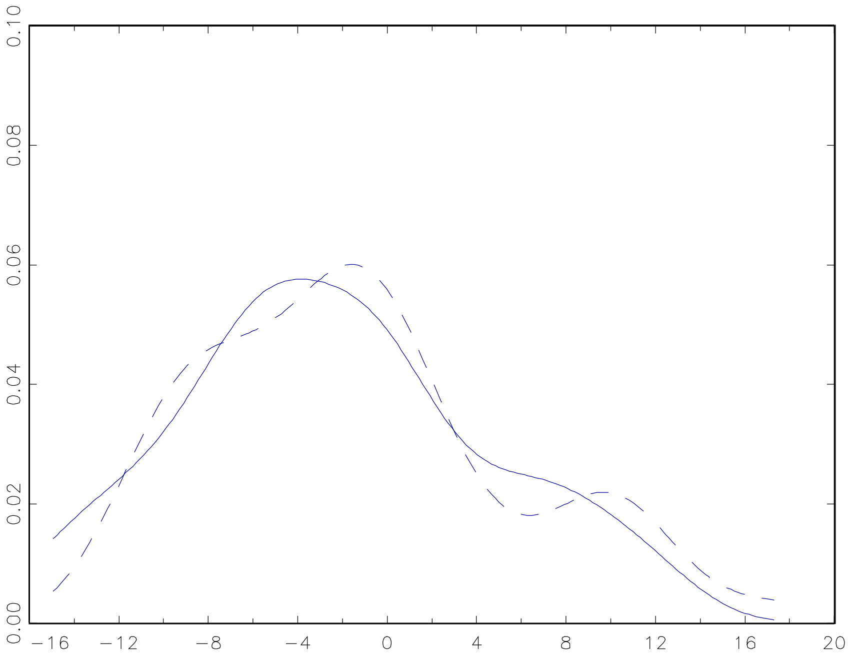
is a very simple comparison of the external shape of the distribution at the two time points. Besides the large dispersion --already detected by figures in Table 1--, at first sight the figure reveals that the shape of the distribution did not experience significant changes. In fact, we have computed an overlapping coefficient that measures to what extent one can accept that both samples were drawn from the same theoretical distribution. A description of this coefficient as well as the critical values to test the null of equal distributions based on bootstrap techniques is in the appendix. The overlapping coefficient for the regional distribution of unemployment rates in 1985 and 1997 takes a value of 0.9325, which is not significantly different from one --equality in the distributions.

However, a closer look at both densities reveals some sort of trend towards the concentration of mass of probability in particular intervals of unemployment rates. The most striking is the consolidation of a peak at very high positive differentials in 1997. Further, an additional one might be in formation on the left of the distribution. Particular overlapping coefficients for three levels of relative unemployment (low, medium and high) confirm this intuition. For the range of large positive differentials, the coefficient equals 0.7632, being significantly below the critical value. So the null of equality for the distribution in both years for that range can be rejected. In the case of the large negative differentials range, the coefficient takes a value (0.8016) close to the critical, which prevents a clear conclusion on the significance of the difference between both distributions in that interval.

Summing up, the shape of the distribution along the period does not seem to have experienced significant changes. If something, there seems to be a tendency towards polarisation in at least two groups of regions regarding unemployment differentials. This polarisation would agree with the phenomenon detected by Overman and Puga (1999) when analysing the distribution of unemployment rates in a set of EU regions.

The estimated density function provides an approximation to the (external) shape of the distribution. However, as Quah (1997) has stressed, movements within the distribution are even more important when analysing the dynamics of a distribution. In our particular study, it would be rather different that provinces with positive differentials at the beginning of the period had most of the probability of remaining in such situation at the end that, on the contrary, they had large probability of transition towards lower differentials. Similarly, regions with low relative unemployment can either have large probability of remaining in such situation or show some propensity to transition to other places in the range of the variable.

Figure 1. Estimated density function of regional differences in unemployment rates.



Notes: Non-parametric estimation by using a Gaussian kernel with bandwith selected as described in Silverman (1986, section 3.4.2).

(Intra)Distribution dynamics can be proxied by the estimation of a stochastic kernel (Stokey and Lucas, 1989) for the regional distribution of unemployment over the period under analysis. This is merely the counterpart of a first order Markov probability of transitions matrix when the number of states tends to infinity, that is to say, when the length of the range of unemployment rates defining each state tends to zero. Thus, the stochastic kernel provides the likelihood of transiting from one place in the distribution of unemployment to the other places[10]. Figure 2 shows the stochastic kernel estimated for the distribution of the unemployment rates in the Spanish provinces over the period 1985-1997 (4-year transitions). The z-axis in the three-dimensional plot measures the density of each pair of points in the x-y space that measure the values of the variable in t and t+4. The lines that run parallel to the t+4 axis measure the probability of transiting from the corresponding point in the t axis to any other point 4 periods ahead. Therefore, the more the mass of probability is concentrated along the positive diagonal, the higher the persistence. On the contrary, a kernel parallel to the t axis will indicate that regardless of the position at time t, all regions tend to show similar unemployment rates at t+4. The two dimensional graph at the upper-right corner is a contour plot of the three-dimensional plot. Lines in this graph connect points at the same height on the three-dimensional plot, that is, points with the same density.

The estimated kernel confirms that the regional distribution of unemployment in the Spanish provinces was characterised by strong persistence[11]. This means that there was strong stability in the ranking of provinces regarding their unemployment rates over the period. Broadly speaking, most of the mass of probability runs along the diagonal defining persistence. If something, shifts from the diagonal indicates the tendency towards the polarisation pointed above. Then, the kernel reproduces the three clusters of regions previously found by the estimated densities[12]. By recursively applying the kernel to the final (as well as to the initial) distribution we can trace out the dynamics of the whole system of economies. Furthermore, when the process is ergodic we can obtain the long-run limit of the distribution from the stochastic kernel. This is a density function that summarises the shape of the distribution in the long-run. Thus, a long-run density function whose mass of probability collapses into a single point (zero in our case) will indicate that the economies under analysis are converging to each other in their unemployment rates. On the contrary, divergence in unemployment rates will translate into a long-run distribution where density is uniformly distributed for the values of the unemployment rate differentials. It is also important to note that long-run polarisation will be confirmed when the ergodic distribution shows *peaks*.

Then, is the dispersion and stability in the unemployment rates differentials in the Spanish provinces a long-term phenomenon? Does the ergodic distribution confirm the formation of clusters of provinces with separate unemployment rates? Figure 3 can shed light on these points. The shape of the long-run distribution is quite similar to the one observed in 1985 and 1997. However, it seems to indicate that the separate cluster at the left end of the distribution that appeared in 1997 is not likely to persist in the future. The peak in the opposite end is not as clear as in the distribution for 1997 but the mass of probability in large values for the differentials is quite evident. So, it is not likely that the large differentials for the high relative unemployment provinces are going to vanish --under the assumption of stability in the dynamics. The value for the overlapping coefficient between the ergodic and the distribution in 1997 confirm these conclusions. For the whole distribution it equals

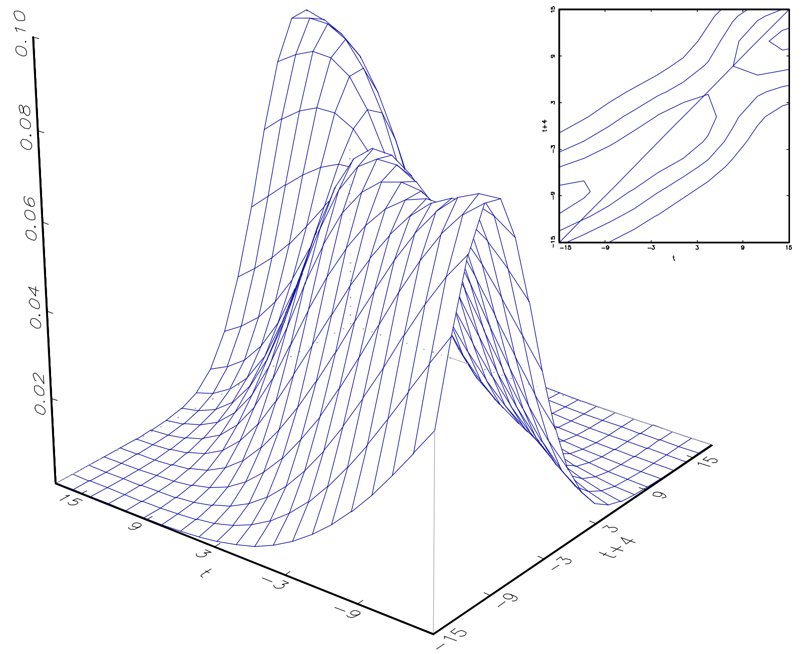
---

[10] See Durlauf and Quah (1999) for a formal definition and some properties of stochastic kernels in the study of distribution dynamics. Since the proposal of their use by Quah (1997), they have been applied by López-Bazo et al (1997), Overman and Puga (1999) and Lamo (2000), among others.

[11] The Gauss code developed by J. Suetrim was used to compute the bivariate density needed to calculate the stochastic kernel. It is available at http://cep.lse.ac.uk/fmg/people/shuetrim/gauss.htm

0.9226, being higher than the critical value, whereas for the three intervals previously defined only in the case of the low relative unemployment we reject the null of equality of the distributions (OVL_{low}=0.8177, OVL_{mid}=0.9327, OVL_{high}=0.9976).

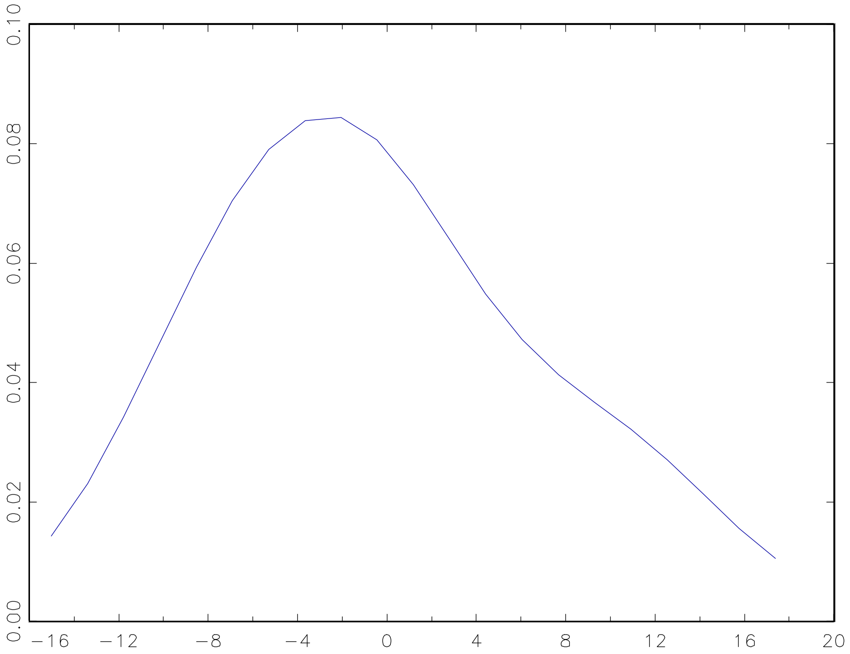
Figure 2. Estimated stochastic kernel for unemployment rates differentials 1985-1997.
4 years transitions.



Notes: Non-parametric estimation by using a Gaussian kernel with bandwith selected as described in Silverman (1986, section 4.3.2).

[12] A similar picture is obtained when the stochastic kernel is estimated considering only the initial and final distributions, that is probability of transitions between 1985 and 1997. This result will be provided from the authors upon request.

Figure 3. Ergodic distribution for unemployment differentials from dynamics between 1985-1997.



Summing up, our results confirm the existence of large differentials in the unemployment rates across the Spanish provinces, whose geographical distribution is characterised by strong persistence. In other words, regions with high relative unemployment did not seem to show any significant trend to change their situation over the period. The same rules for regions with low relative unemployment. Deviations from perfect persistence seem to be only due to some sort of cluster formation in two groups: one with unemployment rates below the average and another with rates far above the average.

The next sections aim at explaining why regions may differ in their unemployment rates and the reasons for the dynamics observed so far. In other words, we will try to assess how much of the dispersion in the geographical distribution of unemployment in Spain can be explained by some of the factors that likely affect unemployment, and to what extent those factors explain the polarisation already observed.

FACTORS EXPLAINING THE GEOGRAPHICAL DISTRIBUTION OF UNEMPLOYMENT

We are interested in analysing the contribution of different variables, which can affect the level of unemployment on a priori grounds, to the dynamics of the geographical distribution of unemployment in Spain. As a consequence, we cannot constrain the study to a traditional regression analysis in which we test if the sign of the parameter for some variables agrees with some theoretical hypothesis, and how much of the variance can be explained with such variables. Instead, we would like to compare the entire observed distribution with the one obtained once the effect of some determinants of unemployment has been removed.

Then, here we build on the analysis proposed by Quah (1997). In a first step we will estimate the effects of some variables than presumably affect unemployment rates by means of a regression analysis. Then, using the tools in the previous section, we will compare the original distribution with the one conditioned to the factors under analysis.

Empirical framework

As previously discussed, unemployment rates differentials in the Spanish regions seem to be only slightly influenced by differences in wages and by migration decisions. Thus, the assumption in this section is that for a given amount of vacancies in a local labour market, unemployment will depend on different types of mismatch, some demographic variables, and neighbourhood effects arising from interactions across labour markets. Concretely, the conditional analysis builds on the so-called Beveridge Curve (see e.g. Blanchard and Diamond, 1989; Pissarides, 1990). Antolin (1994, 1999) and López-Tamayo (2000) have estimated the BC for the Spanish economy. A negative slope for the relationship between unemployment and vacancies is not always clear, but there is strong evidence on the influence of sectoral and skill mismatch, as well as demographic factors affecting participation decisions, in the level of unemployment.

Data availability at the provincial level prevents the inclusion in our analysis of some other factors that have been noted to affect unemployment rates, such as long-term unemployment and unemployment benefits. However, the omission of these factors is supposed not to alter the main results if they have a homogeneous effect in all the provinces, given that we focus on relative unemployment. Furthermore, some of the variables already included may capture at least some of their effects.

The empirical model we will use to obtain the conditional distribution can be expressed as:

$$u_i = f(v_i , Sect_i , Dem_i , h_i , u_{\rho i} ; \beta) + \varepsilon_i \qquad (1)$$

where $f(\cdot)$ is a linear function in all its elements, u_i denotes the difference between the unemployment rate in a province i and the aggregate unemployment rate in Spain, v the same difference in vacancy rates and Sect, Dem and h are the differences in some measures of sectoral composition, demographic variables and a proxy for the skills of the labour force respectively. Finally, $u_{\rho i}$ denotes unemployment differentials in the neighbouring regions, whereas β is a vector of parameters, conformable with the variables in the function, that measure the intensity of the effect of each factor on unemployment differentials. A perturbance (ε) is added to the specification to account for instance for (transitory) different regional responses to shocks.

The share of labour in the agricultural sector over the total labour in the economy and the share in manufactures are the variables proxying for differences in the industrial mix[13]. It is important to note that the Spanish economy experienced an important change in the sectoral composition since the seventies. Labour in agriculture decreased dramatically, particularly in those regions specialised in that activity. Excess of labour in such areas was mostly absorbed by migration to more developed regions up to the early eighties[14]. Afterwards, mechanisms within each region seem to have been the only ones available to balance the labour market. As long as employment in other sector had not absorbed the excess of labour in agriculture, inflows to unemployment or outflows from the labour force should have characterised those economies[15].

On the other hand, shocks in the seventies and early eighties exerted strong influence on the manufacturing sector in Spain. Most of the industrialised regions carried out a profound restructuring process in their manufacturing sector, particularly in the old heavy industries. As a result there was an excess of labour that was not always absorbed by new activities. Hence, it is supposed to exist a positive relationship between the share of labour in these two sectors and the level of unemployment. This would have been probably the case when the shock took place. But if we measure the relationship some time ahead we should expect those economies with the healthiest agricultural and manufacturing sectors having the largest shares while low unemployment. On the contrary, regions that strongly suffered the shocks would have decreased the share of labour in those sectors today, also showing higher unemployment as a result of the job destruction.

As regards the demographic factors, differences across regions in female participation and in the percentage of young population might explain some of the unemployment differentials. Unemployment rates are notably higher for women and especially for people

[13] A description of the variables used in the analysis and their sources is in the Appendix.
[14] As well as migration to other countries.

ageing 16-25. If so, provinces with higher percentage of young population are likely to show positive unemployment differentials, while those with large female participation will tend to show higher unemployment figures. However, in the latter case, the conclusion might be counteracted by the fact that female decisions of participation in the labour market are more related to the current level of unemployment. As a result, female participation decreases in a situation of high unemployment, the opposite being the case when unemployment is low.

Skills, education, and in a broad sense human capital of the labour force is one of the most important factors for an economy to adapt to the ongoing changes in the process of production. Duration in unemployment is inversely related to the level of human capital of a worker. Unemployment rates for workers with upper-grade studies are much lower than for workers with low education (Nickell and Bell, 1996). If the average human capital of the labour force in the Spanish provinces differs, it could explain some of the inequality in the geographical distribution of unemployment.

Finally, we have included the (average) unemployment rate in the adjacent provinces --spatial lag of the unemployment rate-- as an explanatory factor. In this way, we will account for spillover effects in local labour markets, as those described in Burda and Profit (1996). The effects of spillovers across economies on growth have been recently analysed in a bunch of papers (Quah, 1996; Ciccone, 1996; Moreno and Trehan, 1997; López-Bazo et al, 1998). Such interactions across regional labour markets have been shown to explain most of the dynamics in the distribution of the unemployment rates in the EU by Overman and Puga (1999). However, our results indicate that the influence of such effects is lower when the distribution is simultaneously conditioned to other factors in each economy. This advises that behind a significant effect of the spatial lag might be not only some sort of spatial interaction but also some spatially correlated factors omitted in the analysis. Further discussion on this issue will be in the next section.

Conditional distributions

Once an estimation of the parameters in (1) is available we can obtain the distribution of relative unemployment conditional to each one of the factors previously defined, in a given time period. To do so, we first compute the unemployment differentials conditional to each one of the factors. For instance, if we are interested in the effects of human capital in a period t, the conditional distribution will be obtain as the combination of the estimates for the

[15] Marimon and Zilibotti (1998) relate the high Spanish unemployment to the sectoral restructuring in the last decades, particularly to the situation in the agricultural sector.

parameters and the corresponding variables, excluding the human capital variable, plus the vector of residuals. That is to say, we estimate the unemployment differentials in case there is no difference in the level of human capital across provinces, all other things equal.

Then, the density function for the conditional distribution can be estimated and compared to the unconditional one. This way we can assess the contribution of each factor to the particular shape of the real distribution. Besides, we can compute the stochastic kernel relating the unconditional and conditional distributions. It will map the probability of changes in unemployment differentials when the effect of the specific factor is removed. The mass of probability along the positive diagonal in the unconditional-conditional space will reveal no significant changes between both distributions. Thus, we will conclude that the factor under analysis is not important to explain the geographic distribution of unemployment. On the contrary, when it explains most of it, we will observe the kernel parallel to the axis measuring the values of the real unconditional distribution.

The conditioning procedure in this paper is thus in the spirit of Quah (1997), although in our case the conditional distributions are obtained from a more structural model. It differs from the conditioning exercise in Quah (1996) adapted to the analysis of regional unemployment in the EU by Overman and Puga (1999). Those papers compare the distribution of differences between the variable of interest in the regions and the aggregate in the entire economy, with the distribution of differences between the region and the average in a group in which the region is classified. Groups are defined to be either member states in the EU, neighbours, trade partners, economies with the same sectoral composition or similar skill composition of labour.

The latter procedure does not demand the estimation of any parameter to obtain the conditional distribution. But this is at the cost of not allowing the influence of more than one factor to be considered at a time. In contrast, the way we proceed in this paper hypothetically isolates the contribution of each factor, so it is supposed to provide more accurate results on their particular contributions.

Results

We have obtained the distribution of unemployment rates conditional to the same sectoral composition (no differences in shares of labour in agriculture and manufactures), to the same demographic variables (no differences in female participation rates and in the percentage of population ageing 16 to 25), and to similar human capital endowments (no differences in the share of working population that have started at least secondary schooling). Besides, the distribution conditional to have similar neighbours have also been computed, thought deeper comments on its results will be made in the next section. All these distributions have been computed for the initial and final years, 1985 and 1997. This is because even when changes along the period have been shown not to be very important, it is plausible that the effects of the factors had changed.

Table 2 summarises the regression results. There is a set of columns for each one of the years under analysis. For each year there are four estimates. The first two columns collect the OLS estimates of a model with vacancies, shares of labour in agriculture and manufactures, and human capital. The second adds the demographic variables to the model. The last two columns for each year measure the contribution of the neighbouring effects to unemployment rates differentials. Unlike the first of them, in which vacancies and the spatial lag of unemployment are the only regressors, the last one quantifies the effect of the spatial lag when we control for the factors within each province that might influence unemployment rates. The presence of the spatial lag of the endogenous variable causes a problem of endogeneity in the OLS estimator. Maximisation of the likelihood of that model provides with consistent estimates for all the parameters (see Anselin, 1988)[16].

The variables accounting for differences in sectoral composition and human capital in 1985 are highly significant when the demographic variables are not included in the regression. But the intensity of the effect of the share of labour in manufactures and human capital decreases markedly and is only marginally significant when the percentage of young population and female participation are considered as well. From the second column of the table we can observe the negative relationship between vacancies and unemployment hypothesised by the Beveridge curve, although the parameter is estimated with a high dispersion. After controlling for differences in vacancies, the OLS results indicate that the larger the percentage of young population, the higher the unemployment rates differentials. On the contrary, the influence of female participation is negative, although only significant at 10%. As discussed above, this can be due to the fact that the cyclical behaviour of female participation decisions counteract the positive effect of female participation on unemployment.

Differences in the shares of labour in agriculture and manufactures in a region affect negatively to unemployment differentials. This confirms our previous hypothesis: regions in

[16] The likelihood function in our case is shown in the Appendix. The library OPTMUM in the GAUSS system has been used to obtain the ML estimates. Convergence was achieved in a reasonable low number of iterations in all the estimations.

which shocks in the seventies and early eighties stroke deeper, lost employment in those sectors. As a result their shares decreased while largely contributed to increase unemployment that still remained in the mid-eighties. In other words, provinces with (relative) healthy agriculture and manufactures tended to show lower unemployment differentials.

Finally, human capital endowment is negatively related to relative unemployment. An additional point in the percentage of workers with at least secondary schooling in a region relative to the Spanish aggregate decreased by a quarter of point the difference in unemployment rates.

The size of the effects changes a little when spillovers are included in the regression (fourth column in the table). But the most important is that the inclusion of the spatial lag of unemployment causes female participation and human capital to be non-significant. As regarding the parameter of the spatial lag, it shows a moderate value (significant at 10%).

The last four columns collect the results for the same specifications but the variables measured in 1997. The first conclusion when comparing the estimates for the two years is that variables accounting for differences in the sectoral composition and the level of education of the labour force do not seem to contribute in explaining unemployment differentials at the end of the period. These variables do not statistically differ from zero when we account for differences in the demographic factors. The magnitude of their parameters is about half the one observed for 1985 even when demographic factors are not included in the analysis.

On the contrary, the percentage of young population increased its positive influence on the differences in unemployment rates. One additional point of difference in this percentage in a region and the national aggregate translated into more than one additional point in the difference of unemployment rates. This estimate agrees with the evidence on the rates of young unemployment. In the Spanish economy, 36 out of 100 workers ageing less than 25 were unemployed in 1999 --far above the 19% of average for the EU. Differences across Spanish regions are notable: that rate is above 40% in Asturias, Castilla y León, Extremadura and Andalucía, whereas in Aragón and Navarra is below 25%. These figures and our estimates confirm the importance of young unemployment not only to understand the level of unemployment in the Spanish economy but also to explain differentials across areas.

Female participation, that remains significant at 10% even when the spatial lag of the unemployment rates is included, also doubled its parameter in 1997. These results are quite robust to the inclusion of the spillovers across provinces. One interesting feature is that the parameter of the spillover effect almost double its size as well when we include factors within the region in the regression. We will analyse these effects with greater detail in the next section.

From the regression results we could conclude that there was a change in the factors that caused unemployment differentials across the Spanish provinces. But we are also interesting in knowing how much of the dispersion in unemployment differentials were explained by such factors in both periods, and what is the most important, if they may help to explain the polarisation observed in the previous section. To do that, we compute the conditional distributions as described above, compare the estimated densities for those distributions with the real ones and estimate the stochastic kernel mapping their joint probabilities.

Results for the distribution of unemployment rates conditioned to (no differences in) the sectoral composition, human capital and demographic factors are shown in Figures 4 to 6. Plots in the first column are the densities (up) and the kernel (bottom) for 1985, while those for 1997 are in the second column. The continuos line represents the density for the real (unconditional) distribution, while the dotted line is for the conditional one. In the stochastic kernel, the x-axis measures the values of the real distribution and the y-axis the conditional one.

It can be observed how when removing the effect of the differences in the sectoral composition (figure 4), the distribution at the beginning of the period is much more concentrated around zero (no differences in unemployment rates). That is, differences in such magnitudes explain an important amount of the dispersion of the distribution in 1985. However, the shape of the right tail is quite similar in both distributions. This is confirmed by the value of the overlapping coefficient, that for the entire range is 0.810, below the critical value, and for the negative and around zero differentials (0.681 and 0.320) is far below the critical values, whereas for the positive differentials range equals 0.829.

The same picture is reflected in the stochastic kernel. The mass of probability in the range of low relative unemployment (negative differentials) departs from the diagonal. Actually, for all the negative values it almost runs parallel to the axis of the unconditional distribution, while for the positive differentials it returns to the diagonal. This means that sectoral differences were highly responsable for the low relative levels of unemployment in some Spanish provinces in 1985.

Table 2. Regression results.

	1985				1997			
	OLS	**OLS**	**ML**	**ML**	**OLS**	**OLS**	**ML**	**ML**
Constant	-0.856 (0.875)	-0.821 (0.778)	-0.871 (0.757)	-0.297 (0.748)	-1.249 (0.880)	-0.782 (0.763)	-0.574 (0.641)	-0.211 (0.672)
Vacancies	-180.209 (289.572)	-129.526 (252.122)	350.719 (272.080)	-71.350 (221.584)	473.761 (104.313)	260.969 (101.262)	279.460 (91.075)	211.371 (90.161)
Young		0.888 (0.225)		0.711 (0.216)		1.406 (0.317)		1.022 (0.291)
Female participation		-0.131 (0.093)		-0.095 (0.084)		-0.254 (0.164)		-0.210 (0.136)
Labour in agriculture	-0.752 (0.105)	-0.476 (0.110)		-0.424 (0.103)	-0.392 (0.152)	-0.162 (0.139)		-0.142 (0.110)
Labour in manufacturing	-0.456 (0.137)	-0.225 (0.128)		-0.204 (0.112)	-0.288 (0.114)	-0.132 (0.104)		-0.053 (0.093)
Human capital	-0.392 (0.152)	-0.241 (0.135)		-0.167 (0.130)	-0.178 (0.163)	-0.151 (0.139)		-0.092 (0.115)
Spatial lag of unemployment			0.697 (0.108)	0.218 (0.148)			0.631 (0.108)	0.369 (0.147)
Adjusted R^2	0.513	0.645			0.415	0.581		
LnL			-154.542	-136.27			-147.322	-140.066
F (p:)	13.898 (0.000)	15.849 (0.000)			9.701 (0.000)	12.33 (0.000)		

Notes: Dependent variable is the difference between unemployment rates in each province and the nation-wide aggregate. All the regressors are also in deviations to the value for the Spanish aggregate. Robust standard errors in brackets. LnL is the maximum of the log-likelihood. F(p:) denotes the joint signicance test and the associated p-value respectively. All the estimates included the observations of the 50 Spanish provinces.

The same does not apply to the distribution at the end of the period. In accordance with the regression results, conditional to the sectoral composition the distribution remains unaltered in 1997 (the OVL coefficient for the entire range and the different intervals is closer enough to one --OVL=0.960, OVL_{low}=0.906, OVL_{mid}=0.942, OVL_{high}=0.817-- not to reject the equality of both distributions).

As regards the distribution conditional to human capital, it is clear from figure 5 how differences in education of the labour force exert small influence in the geographical distribution of unemployment. The most significant feature is its contribution to exacerbate differentials for provinces with very high unemployment in 1985. However, even for the range of high positive differentials, the overlapping coefficient does not reject that both distributions are similar for that interval.

Finally, the comparison of densities and the stochastic kernels when the distribution is conditional to the percentage of young population and to female participation (figure 6) confirms the important contribution of such factors to explain the spatial distribution of unemployment in both years. The influence of demographic factors seems to increase at the end of the period, especially when explaining the polarisation observed in the real distribution. While the overlapping coefficients for the extreme intervals are around the critical values --preventing from a clear conclusion-- in 1985, they clearly reject equality in those ranges in 1997. However, the conditioning does not remove the peak at positive differentials, but simply move it closer to the point of no differentials. Therefore, we cannot but conclude that the polarisation should be due to factors other than those traditionally considered to explain unemployment. In the next section we try to see to what extent this can be explained by dependence across the space.

Figure 4. Conditioning to (no differences in) sectoral composition.

— real distribution --- conditional distribution

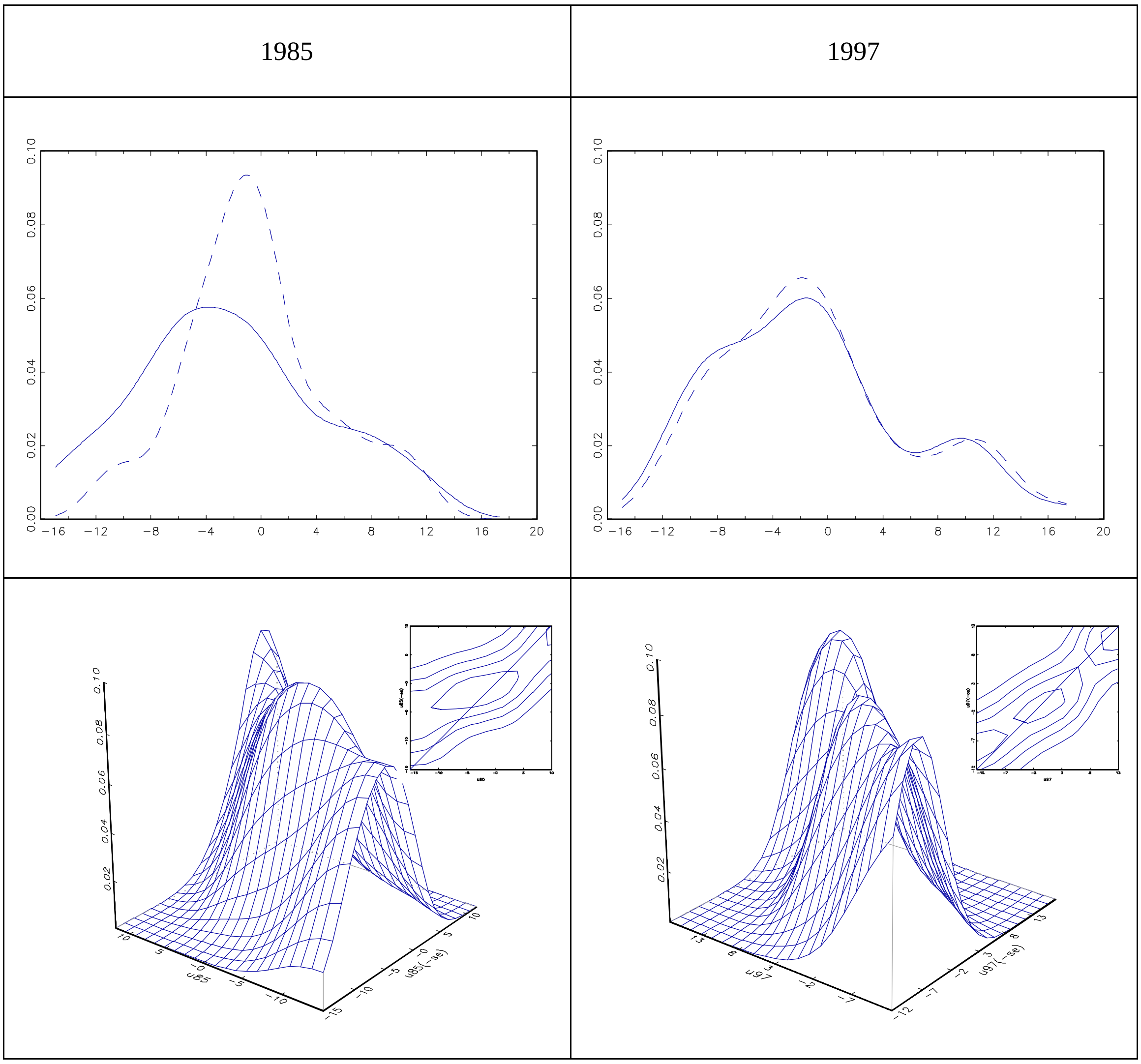


Figure 5. Conditioning to (no differences in) human capital.

— real distribution --- conditional distribution

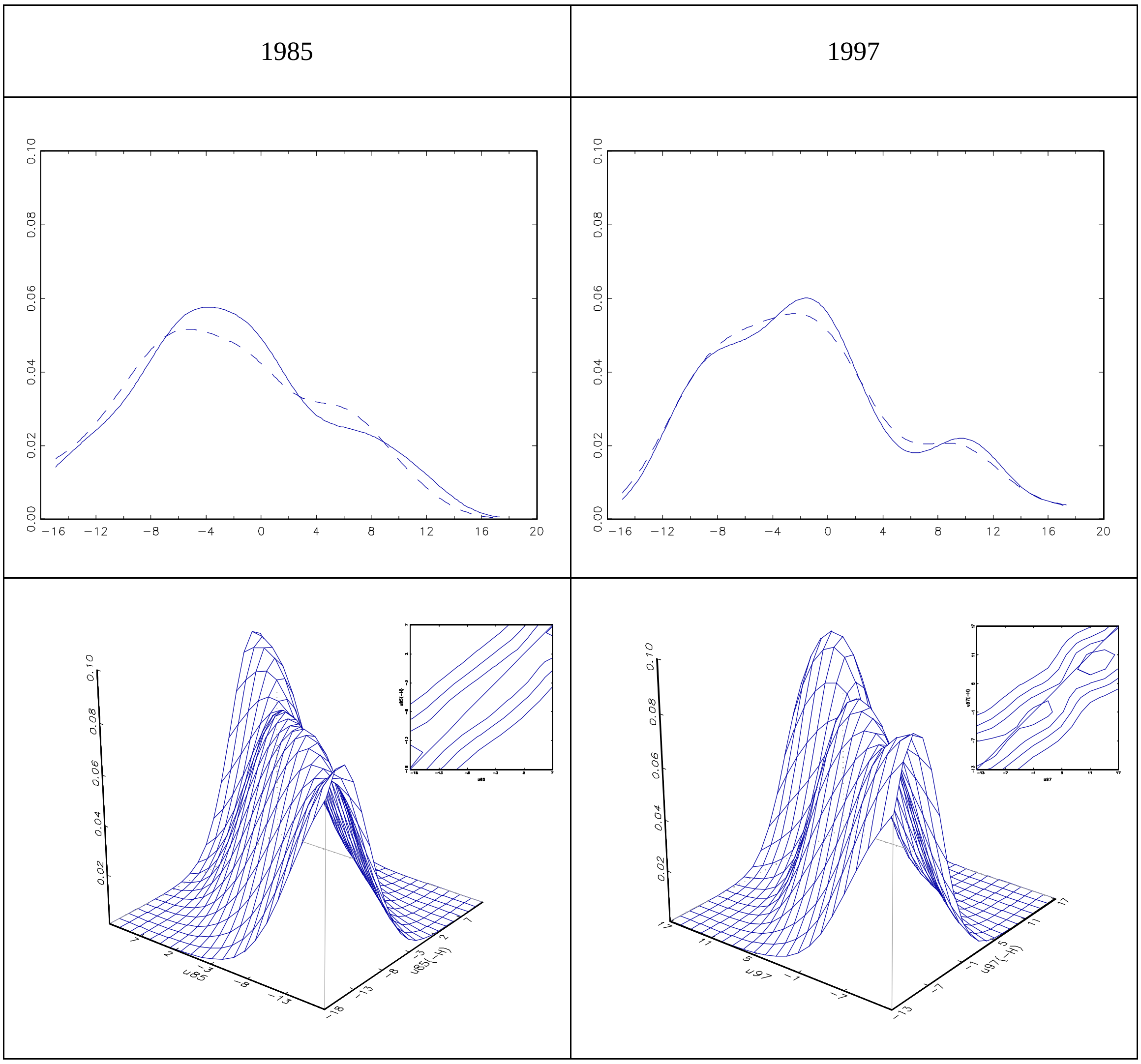
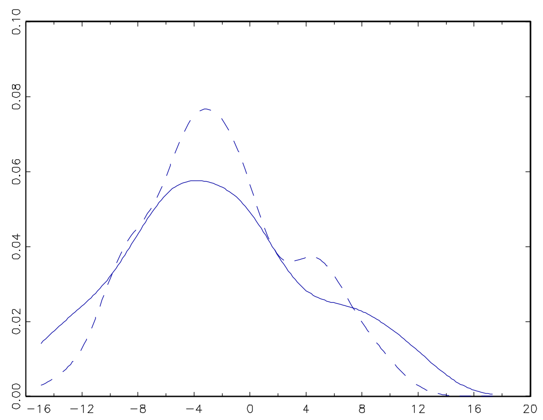
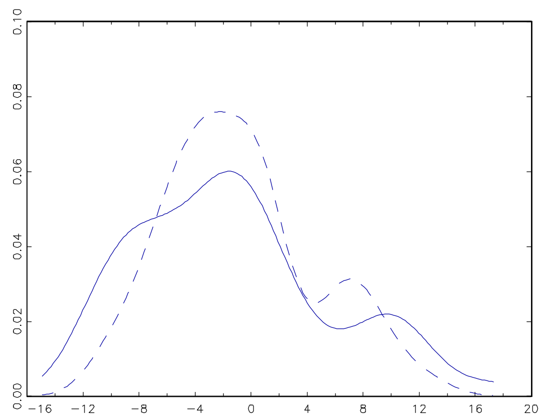
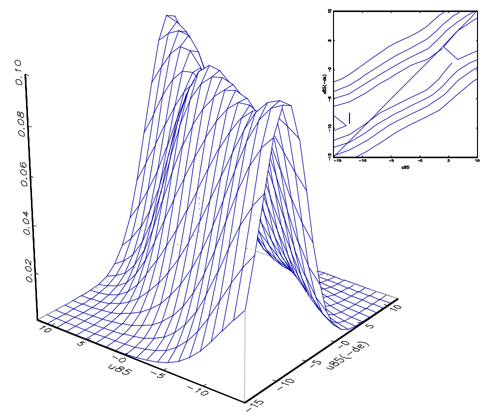
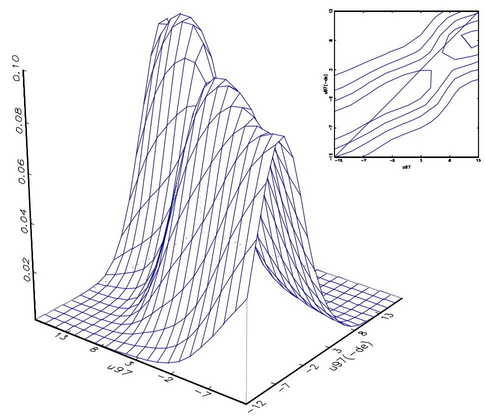


Figure 6. Conditioning to (no differences in) demographic factors.

— real distribution --- conditional distribution

1985	1997
	
	

NEIGHBOURING EFFECTS ON UNEMPLOYMENT

"Spain's cooking can be considered by way of the different regions, although there are similarities between neighbouring areas" Iberia Airline Magazine, December 1999.

Could the geographical location of the economies induce their level of development, economic activity and the other aggregate economic magnitudes, among them unemployment? As mentioned above, some empirical evidence has recently been provided pointing to the influence of spillovers across economies on growth and location of activity (Quah, 1996; Ciccone, 1996; Moreno and Treham, 1997; López-Bazo et al, 1998; Rey and Montouri, 1999)[17]. If the geographical scope of the process of technological diffusion, agglomeration economies or any other force that may cause the activity to cluster is not restricted to the boundaries of a single economy, interdependencies might explain why spatial groups of economies sharing similar economic characteristics are usually observed[18].

In the case of labour markets, Burda and Profit (1996) have proposed some sources for interactions across labour areas that, in brief, are due to workers in a region willing to fill vacancies in other regions and firms looking for workers outside the regions in which they are located. Their empirical evidence reveals that unemployment and vacancies in nearby regions do matter to explain matching in local labour markets in the Czech Republic. Similarly, Burgess and Profit (1998) for the travel-to-work areas in Britain and López-Tamayo (2000) have provided related empirical evidence.

From another perspective, Bronars and Jansen (1987) and Molho (1995) report the significance of spatial spillovers in the process of adjustments of unemployment differentials to local shocks in UK and in the US respectively.

Finally, in an empirical framework more related to the one in this paper, Overman and Puga (1999) report that the distribution of the unemployment rates in the EU regions relative to the rates in the neighbouring regions is much more concentrated than the distribution relative to the EU-wide average rates. Contiguity seems to be more important to understand the dynamics within the regional distribution of unemployment in the EU than membership to a given country, similarity in the sectoral composition or

[17] At the individual level such interactions have also shown to be important. When making decisions, agents tend to make use of decisions others made in similar situations. See for instance Bénabou (1996) and Kapteyn et al (1997).

in the skills of the labour force. However, their conditional analysis does not allow isolating neighbouring effects from factors within each economy. The procedure described in the previous section does permit so for our sample of regions.

A simple look at figures of unemployment rate differentials in the Spanish provinces reveals that they are not randomly distributed in the space[19]. Provinces with high unemployment rates are likely to be surrounded by provinces with high rates (particularly provinces in the South), the opposite being the case for those with low unemployment (notably in the North-East of Spain). Actually, spatial dependence statistics for the distribution of unemployment rates do strongly reject the null of no dependence. The values for those tests also point to increasing spatial dependence between 1985 and 1997.

As described in the previous section, the last two columns in the panels for each one of the years under analysis in table 2 incorporate the spatial lag of the unemployment differentials. The first of them only include this variable, besides vacancies. The other simultaneously considers the factors within each province that might affect unemployment. It is quite clear how the omission of the factors within provinces makes the estimate of the parameter of the spatial lag to increase markedly. This is especially the case in 1985. In fact in the full specification, the spatial lag is only significant at 10% that year.

The densities and kernels for the distributions conditional to (the absence of differences in) the spatial effects when the other factors are discarded are in figure 7.
It is evident how when unemployment differentials are computed under the assumption that all provinces have *similar neighbours*, the distribution concentrates around the point of no differences in both years. Besides, spatial effects seem to explain most of the polarisation observed in the real distribution. In other words, most of the regions in the clusters of low and high relative unemployment are located together in the space.

This result agrees with the one in Overman and Puga (1999) for the EU regions. But how robust are they to the inclusion of idiosyncratic regional factors? That is, would it be possible that most of the effects captured by unemployment in the neighbours were explained because factors causing unemployment do not differ markedly across contiguous regions?

---

[18] López-Bazo et al (1999) provide evidence of this for the GDP per capita and for the productivity levels in the EU regions, while Rey and Montoury (1999) do it for income in the US States, by applying spatial statistics.
[19] Obviously they are not homogeneously either.

Figure 7. Conditioning to (no differences in) neighbouring effects (when considered in isolation).

— real distribution --- conditional distribution

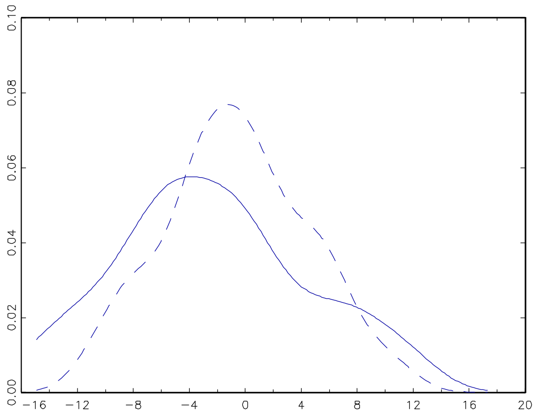
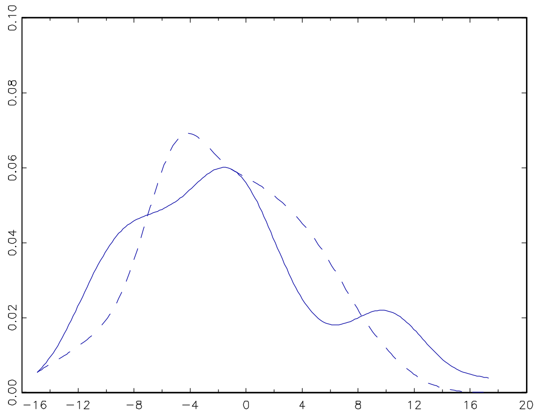
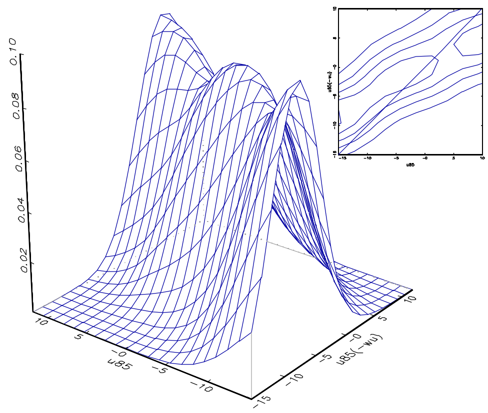
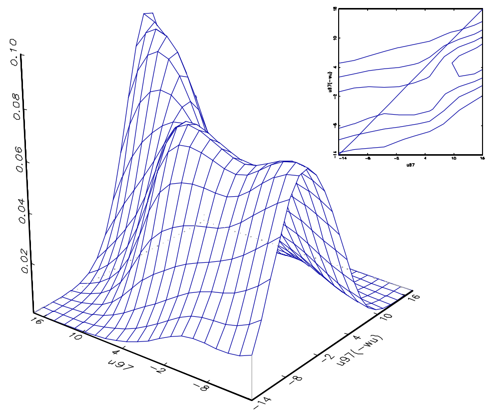
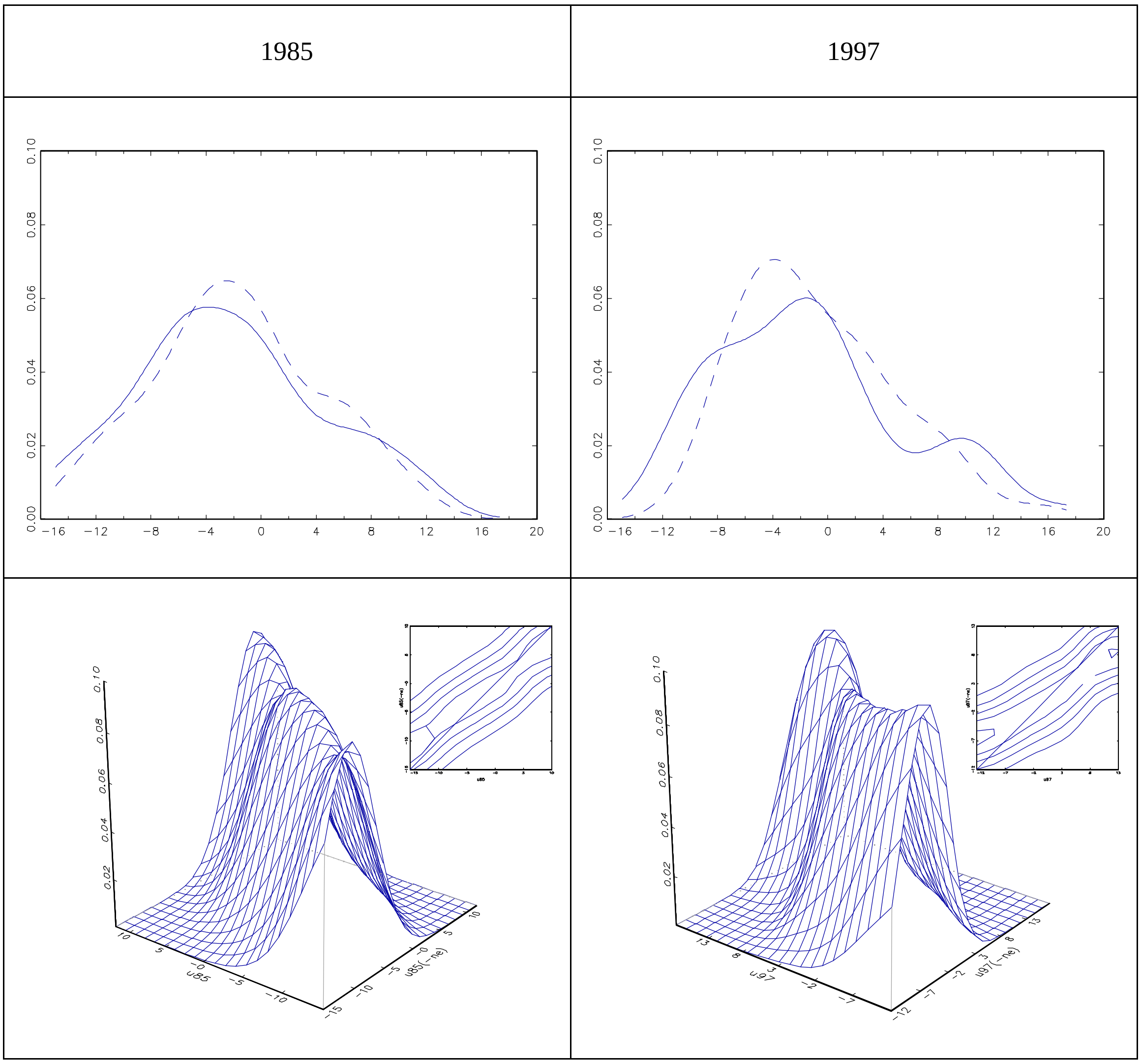
1985	1997
	
	

Figure 8. Conditioning to (no differences in) neighbouring effects (when factors within the region are included).

— real distribution --- conditional distribution



To some extent the evidence in figure 8 indicates that the answer is yes. When we obtain the distribution conditional to the unemployment in the neighbours, but having included the other factors, the contribution of the spatial effects decreases, especially in 1985. The overlapping coefficient is only (slightly) below its critical level for the medium interval that year, but for the two tails and for the entire distribution we cannot reject that the unconditional and the conditional distributions do not differ.

At the end of the period, however, spatial spillovers still can explain polarisation. In the conditional distribution, the mass of probability at the negative differentials disappears, whereas the distribution no longer shows the discontinuity at the very high positive differentials. This could be related to the decrease in the explanation attributable to the traditional factors considered in the analysis at the end of the period. If this was the case, interactions across regional labour markets could explain some of the differentials. But also factors within each economy others than the ones considered here, that probably do not differ a lot across nearby regions, could be playing an increasing important role.

FINAL REMARKS

By studying the dynamics of the geographical distribution of the unemployment rates in Spain we have been able to assess the magnitude of regional differentials and how persistent those differentials are over time. Interestingly, the observation of the (external) shape of the distribution has revealed some trend toward the polarisation of unemployment rates into two groups of regions, one of them with extremely high unemployment differentials. Additionally, this feature could run parallel to a process of continuous decrease in aggregate unemployment rates.

From the conditioning exercise we cannot but conclude that the geographical distribution of unemployment has become less dependent on spatial differences in the sectoral composition over time. It is likely that the shocks that caused a restructuring process in the seventies and early eighties still exerted some influence at the beginning of the period considered in our analysis. But they might have been completely eliminated at the end of the nineties. In addition the industrial mix in the Spanish regions showed an important process of convergence, such that regional differences in the share of each sector are much lower nowadays than some years ago.

Unlike some evidence in the literature on the negative relationship between the aggregate unemployment rates and education, differences in the degree of education of the labour force in the Spanish provinces do not help to explain much of the dispersion in unemployment rates. This seems to be particularly the case at the end of the period. In this sense, it should be remembered that the level of education of the labour force experienced outstanding increases in all the Spanish regions in the last decades. The percentage of the labour force that at least started secondary schooling for the entire Spanish economy in 1980 was 26.2%, whereas in 1997 it increased up to 67.0%. In 1980 the percentage for the region with the lowest average education was 10.3% and the one for the province with the highest 39.6%. In 1997 the absolute difference in those percentages was similar, but then the value for the province with the lowest endowment was around the 50%. Therefore, it does not seem likely that further increases in education alone could contribute to decrease disparities. Besides, mismatch between educational supply and labour demand is some Spanish regions is particularly striking. As a result, the level of education of the unemployed is on average above the one of the employed (Rodríguez-Pose, 1996).

Factors that seem to contribute most to the explanation of the spatial differentials in unemployment rates are the demographic ones. When removing the effects on the differences in female participation and especially in the percentage of young population unemployment rates are much more evenly distributed across the Spanish provinces. This was so in the mid-eighties and even more markedly at the end of the nineties. However, they cannot completely explain the cluster of regions with high positive differentials.

We have shown that we can account for at least some of the polarisation by considering spatial effects. The results show that unemployment differentials are spatially dependent, in such a way that when considering unemployment rates in the contiguous regions as well as the idiosyncratic factors, most of the features in the geographical distribution of unemployment are explained. The spatial effects might be proxying different types of interactions across local labour markets but also the omission of factors within each region that do not differ significantly across nearby regions. Further analysis would be required to shed light on this issue.

Finally, we have not studied the role of regional differences in the profitability of private investments and in their capability to attract new activities. The influence of those factors, as well as the possibility of differences in the regional responses to the recent policies aiming at increasing flexibility in the labour market, on the geographical distribution of unemployment are in our future research agenda.

ACKNOWLEDGMENTS

A very preliminary draft of this paper was presented under the title "Regional differences of unemployment in Spain" at the 36th Congress of the European Regional Science Association (Zurich, 1996). The authors want to thank suggestions in the different phases of the realisation of this paper by A. Rodríguez-Pose and J. López-Tamayo (who kindly provided us with data on vacancies). We also benefited from discussions with D. Puga during a hot summer afternoon in Barcelona. Financial support from the DGICYT SEC99-0700 is gratefully acknowledged.

REFERENCES

Antolín P. (1994) Los flujos de trabajadores en el mercado de trabajo español. *Papeles de Economía.* 72, 154-176.

Antolín P. (1999) La curva de Beveridge y las disparidades regionales en España. Papeles de Economía Española, 80, 69-78.

Anselin (1988) *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Bean C. R. (1994) European Unemployment: a survey. *Journal of Economic Literature*, 32, 573-619.

Bénabou R. (1996) Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance. *American Economic Review*, 86, 584-609.

Bentolila S. and Dolado J. J. (1991) Mismatch and internal migration in Spain, 1962-86. In Padoa-Schioppa (ed) *Mismatch and labour mobility*, Cambridge Univ. Press, Cambridge.

Bertola G. and Ichino A. (1996) Wage inequality and unemployment: US versus Europe. *CEPR Discussion Paper* # 1186, London.

Bianchi M. (1997) Testing for convergence: evidence from non-parametric multimodality tests. *Journal of Applied Econometrics*, 12, 393-409.

Blanchard O. and Diamond, P. (1989) The Beveridge Curve. *Brookings Papers on Economic Activity*, 1, 1-60.

Blanchard O. and Jimeno J.F. (1995) Structural unemployment: Spain versus Portugal. *American Economic Review*, 85 (2) 212-218.

Blanchard O. and Katz L. (1992) Regional evolutions. *Brookings Papers on Economic Activity*, 1, 1-75.

Bradley E. L. Jr (1985) Overlaping Coefficient. In Kotz S. and Johnson N.L. (eds.), *Encyclopedia of Statisitical Sciencies*, (6), 546-547.

Bronars S. G. and Jansen D. W. (1987) The geographic distribution of unemployment rates in the US. A spatial-time series analysis. *Journal of Econometrics*, 36, 251-279.

Burda M. C. and Profit S. (1996) Matching across space: evidence on mobility in the Czech Republic. *Labour Economics*, 3, 255-278.

Burguess S. and Profit S. (1998) Externalities in the matching of workers and firms in Britain. CEPR Discussion Paper # 1854. London.

Ciccone A. (1996) Dynamic externalities, spatial technology diffusion, and economic growth. University of California at Berkeley and University Pompeu Fabra. Mimeo.

de la Fuente A. (1999) La dinámica territorial de la población española: un panorama y algunos resultados provisionales. *Revista de Economía Aplicada*, 7, 53-108.

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

Dolado J. J. and Jimeno J. F. (1997) The causes of Spanish unemployment: A structural VAR approach. *European Economic Review*, 41, 1281-1307.

Durlauf S. N. and Quah D. (1999) The new empirics of economic growth. In John B. Taylor and Michael Woodford (eds.), *Handbook of Macroeconomics*, Ch. 4, pp 231-304, North-Holland Elsevier Science.

Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap.* New York. Chapman and Hall.

Esteban J.M. (1999) Un análisis de las desigualdades interregionales en Europa: la década de los ochenta. In Castells A. and Bosch N. (eds) *Desequilibrios territoriales en España y Europa*. Ariel Economía. Barcelona.

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

Hall R. E. (1972) Turnover in the labor force. *Brookings Papers on Economic Activity*, III, 709-764

Inman H.F. and Bradley E. L., Jr. (1989) The Overlaping Coefficient as a Measure of Agreement Between Two Probability Distributions and Point Estimation of The Overlap of Two Normal Densities *Comunications in Statistics – Theory and Methodology*, 18, 3852-3874.

Jimeno J.F. and Bentolila S. (1998) Regional unemployment persistence (Spain, 1976-1994). *Labour Economics*, 5 25-51.

Kapteyn A., van de Geer S., van de Stadt H. and Wansbeek T. (1997) Interdependent preferences: an econometric analysis. *Journal of Applied Econometrics*, 12, 665-686.

Lamo A. (2000) On convergence empirics: some evidence for Spanish regions. *Investigaciones Económicas*, forthcoming.

Lazar F. (1977) Regional unemployment rate disparities in Canada: some possible explanations. Canadian Journal of Economics, 10, 112-129.

López-Bazo E., Vayá E., Mora A. J. and Suriñach J. (1997) Convergencia regional en la Unió Europea ante el nuevo entorno económico. *ICE, Revista de Economía*, 762, 25-41.

López-Bazo E., Vayá E. Moreno, R. and Suriñach J. (1998) Grow neighbour, grow, grow... Neighbour be good! University of Barcelona. Mimeo.

López-Bazo E., Vayá E., Mora A. J. and Suriñach J. (1999) Regional economic dynamics and convergence in the EU. The Annals of Regional Science, 33, 343-370.

López-Tamayo J. (2000) Ensayos sobre la función de emparejamientos en el mercado laboral español. Unpublished PhD thesis, University of Barcelona.

Lorente J. R. (1992) La dispersión geográfica de los salarios. In *Síntesis Mensual de Indicadores Económicos*, Ministerio de Economía y Hacienda, Madrid.

Magrini S. (1999) The evolution of income disparities among the regions of the European Union, *Regional Science and Urban Economics*, 29, 257-281.

Marimon R. and Zilibotti F. (1998) "Actual" versus "virtual" employment in Europe. Is Spain different? *European Economic Review*, 42, 123-153.

Martin R. (1997) Regional Unemployment Disparities and their Dynamics, *Regional Studies*, 31, 237-252.

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

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

Molho I. (1995) Spatial autocorrelation in British unemployment. *Journal of Regional Science*, 35, 641-658.

Moreno R. and Trehan B. (1997) Location and the growth of nations. *Journal of Economic Growth*, 2, 399-418.

Nickell S. and Bell B. (1996) Changes in the distribution of wages and unemployment in the OECD countries. *American Economic Review*, 86 (2) 302-308.

OECD (1989) Regional unemployment in OECD countries. In *Employment Outlook*, pp 95-131. Paris.

Overman H.G. and Puga D. (1999) Unemployment clusters across European regions and Countries, *CEP Discussion Paper* # 434, London.

Pérez , P. and Serrano, L. (1998) *Capital Humano, crecimiento económico y desarrollo en España (1964-1997).* Fundació Bancaixa. Valencia.

Pissarides, C. (1990) *Equilibrium unemployment theory.* Basil Blackwell Ltd. Oxford.

Pissarides C. and McMaster I. (1990) Regional migration, wages and unemployment: empirical evidence and implications for policy. *Oxford Economic Papers*, 42, 812-831.

Pissarides C. and Wadsworth J. (1989) Unemployment and the interregional mobility of labour. The Economic Journal, 99, 739-756.

Quah D. (1996) Regional convergence clusters across Europe. *European Economic Review*, 40, 951-985.

Quah D. (1997) Empirics for growth and distribution: Stratification, polarization, and convergence clubs. *Journal of Economic Growth*, 2, 27-59.

Rey S. and Montouri B. D. (1999) US regional income convergence: a spatial econometric perspective. *Regional Studies*, 33: 145-156.

Rodríguez-Pose A. (1996) Educación superior, mercado de trabajo y crecimiento económico en una España dispar. *Estudios Económicos*, 3, 45-79.

Røed K. (1997) Hysteresis in unemployment, *Journal of Economic Surveys*, 11, 389-418.

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

Schweitzer, M.E. and Dupuy, M. (1996) “Sectoral Wage convergence: a nonparametric distributional analysis” Working Paper 9611 of the Federal Reserve Bank of Cleveland.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York

Stokey N. L.and Lucas R. E. Jr (1989) *Recursive Methods in Economic Dynamics*. Harvard Univ. Press, Cambridge MA.

APPENDIX

Units of analysis

Our definition of region corresponds to the level three of the Nomenclature of Territorial Units for Statistics (NUTS) of EUROSTAT. They are the 50 Spanish provinces. The average surface of a representative province is 10.120 Km^2, while the minimum corresponds to Guipúzcoa, with 1.980 Km^2 and the maximum to Badajoz with 21.766 Km^2. Most of commuting is within these territorial units, so we can say that they approximately define integrated labour markets. The size of the representative (average) province, as measured by the labour force, was 571654 workers in 1985 and 644879 in 1997. The larger market belongs to Madrid (3569900 in 1985 and 4158150 in 1997) and the smaller to Soria (79040 in 1985 and 77880 in 1997) in both periods.

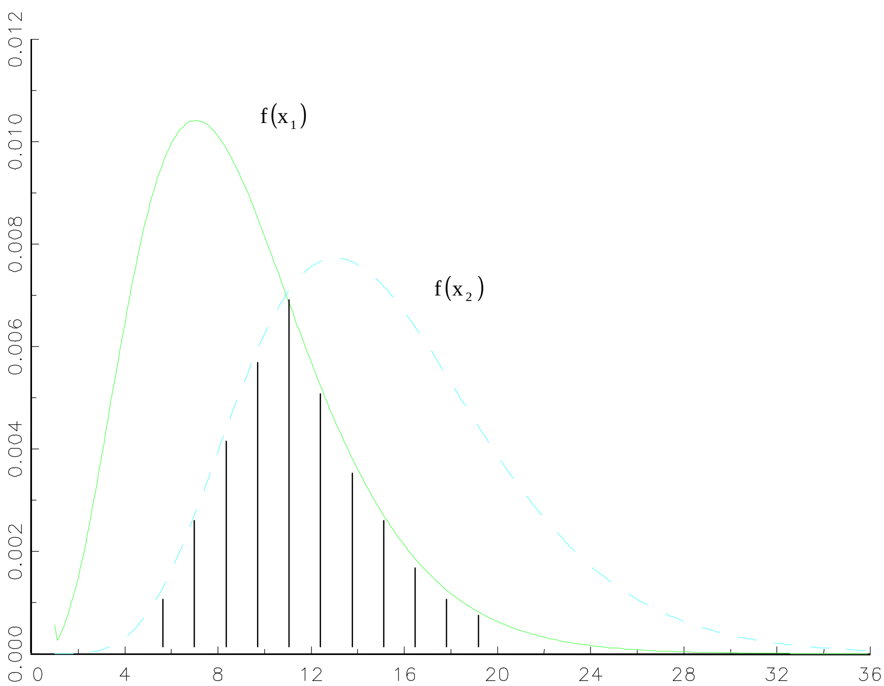
Data description and sources

Variable	Definition	Source
Unemployment Rate (u_i)	$u_i = \frac{U_i}{L_i} * 100$ Where: U_i : Total unemployed force in region i L_i : Total labour force in region i	Labour Force Survey (EPA) from the Spanish Statistical Office (INE) http://www.ine.es.
Vacancy rate (v_i)	$v_i = \frac{V_i}{L_i} * 100$ Where: V_i : Corrected vacancies in region i. $V_i = \left[1 + k \frac{OUT_i^N}{OUT_i^U}\right] V_i^A$ k: Relative efficiency factor of the INEM respect to the market. Set to 0.25 OUT_i^N : Outflows from unemployment unnamed in region i. OUT_i^U : Outflows from unemployment named in region i V_i^A : Administrative vacancies in region i L_i : Total labour force in region i	Elaborated as proposed by Antolin (1996), using data from the Spanish National Employment Office (INEM)
Human Capital (h_i)	$h_i = \frac{H_i}{L_i} * 100$ Where: H_i : Labour force that has at least started secondary schooling in region i L_i : Total labour force in region i	From Pérez and Serrano (1998), taking as primary source the EPA

Labour in Agriculture $(agri_i)$	$agri_i = \frac{AGRI_i}{N_i} * 100$ Where: $AGRI_i$: Employment in agriculture in region i N_i : Total employment in region i	EPA
Labour in Manufacturing (ind_i)	$ind_i = \frac{IND_t}{N_i} * 100$ Where: IND_i : Employment in manufacturing in region i N_i : Total employment in region i	EPA
Female participation (fem_i)	$fem_{it} = \frac{FEML_i}{FEM16-65_i} * 100$ Where: $FEML_i$: Female labour force in region i $FEM16-65_i$: Females in working age in region i	EPA
Share Young (you_i)	$you_i = \frac{YOU_i}{N16-65_i} * 100$ Where: YOU_i : Population from 16 to 25 years old in region i $N16-65_i$: Population in working age in region i	EPA

Overlapping Coefficient

Bradley (1985) and Inman and Bradley (1989) proposed an overlapping coefficient (OVL) as an intuitive measure of the similarity between two probability distributions. In our case we use the OVL to compare the frequencies throughout the range of a variable for two samples. The idea behind the OVL can be summarised in the following figure:



The OVL is the area where the densities of the two distributions overlap when are plotted in the same axes.

The expression for this coefficient in the discrete case is the following:

$$OVL = \sum_x \min[f(x_1), f(x_2)]$$

$$OVL \in [0,1]$$

where $f(x_1)$ and $f(x_2)$ are the empirical density functions. In the case of continuos distributions summation is replaced by integration. As closer the OVL to 1 the more similar the two density function of the samples we are comparing. In the two extreme opposite cases, the unit value for the OVL means that the two density functions are exactly equal, whereas a null value indicates the absence of overlapping in the density function at any point in the range of the variable.

If we wish to assess the contribution of the different individuals in the sample to differences in the distributions, it is possible to compute the OVL for different intervals of the total range of the variable, using the following expression:

$$OVL_\alpha = \frac{\sum_{x \in \alpha} \min[f(x_1), f(x_2)]}{\sum_{x \in \alpha} \max[f(x_1), f(x_2)]} \in [0,1]$$

where α denotes a specific interval.

We have computed the OVL_α for tree different intervals of the unemployment rates differentials (α=low, medium and high). OVL_{low} considers values from the minimum to the average minus one standard deviation of the unemployment rate, OVL_{high} goes from the average plus one standard deviation to the maximum of the unemployment rate. OVL_{med} measures the discrepancy of the distribution in between.

The statistical properties of the OVL coefficients depend on the ones of the data under analysis. So the way to approach this issue is via simulation. Additionally, the OVL is a biased statistic, because any sampling variation in the densities of two samples obtained from the same population causes the OVL being strictly less than one.

We use the bootstrap method to obtain the mean and the variance of the OVL. We do this both from resampling the original data and from a simulated sample of the same size from a $N(\bar{x}_i, s_i)$(i=85, 97). The number of replications used is m=10.000. The process followed to obtained the results reported in the tables bellow is the following:

a- Resample with replacement from the original data and the $N(\bar{x}_i, s_i)$, with sample equal to the number of observations in the original data (n=50).

b- Compute the density function of the resample data.

c. Calculate the OVL and store the results (OVL_j, j=1,2...,m).

d. Repeat steps a to c, until 10.000 replications.

e. Calculate the mean and variance of OVL using:

$$E[OVL] = \frac{1}{m}\sum_{i=1}^{m} OVL_j$$

$$VAR[OVL] = \frac{m}{1-m}\left[\frac{\sum_{i=1}^{m} OVL_j^2}{m} - \left(\frac{\sum_{i=1}^{m} OVL_j}{m}\right)^2\right]$$

Table A1. Expected Value and variance of the OVL by bootstrapping (10000 replications)

	Resample over original data		Resample over $N(\bar{x}_i, s_i)$ i=85, 97	
	E[OVL]	VAR[OVL]	E[OVL]	VAR[OVL]
85	0.9181	0.0009	0.9061	0.0012
97	0.9161	0.0009	0.9036	0.0011

Table A2. Expected Value and Variance of OVL_α by bootstrapping (10000 replications)

	Resample over original data		Resample over $N(\bar{x}_i, s_i)$ i=85, 97	
	E[OVL]	VAR[OVL]	E[OVL]	VAR[OVL]
85-Low	0.7892	0.0122	0.8188	0.0104
85-Mid	0.8728	0.0028	0.8352	0.0034
85-High	0.8093	0.0106	0.8121	0.0101
97-Low	0.8405	0.0106	0.8271	0.0112
97-Mid	0.8742	0.0027	0.8357	0.0035
97-High	0.7567	0.0139	0.7854	0.0090

We use these tabulated values in tables A1 and A2 to construct a sort of confidence interval to test the hypothesis of equality of two distributions. The rule of thumb is to reject the hypothesis of similar distributions if the value estimated for the OVL is lower than the expected value for the OVL in each case minus two times the standard deviation. Then, we reject the null when the overlap is even lower that the one we could expect allowing for sample deviations considering the size of our sample. On the contrary, when the OVL is closer to one than the critical value we can put some confidence on the assumption of equality.

Maximum likelihood estimation of the model with spatial effects

Equation (1) can be expressed as follows:

$$u = \alpha_1 + \alpha_2 v + \alpha_3 \text{Sect} + \alpha_4 \text{Dem} + \alpha_5 h + \rho Wu + \varepsilon$$

where u, v, Sect, Dem, h and ε are (Nx1) vectors of the variables and the perturbance. W is a (NxN) matrix of weights, were each of its elements w_{ij} reflects the interaction between region i and region j. In the case of physical contiguity w_{ij}=1 if region i and region j share common barrier and 0 otherwise. Finally this matrix is row standardised. Therefore Wu is the average of the unemployment rates in the contiguous regions.

The (log) likelihood function for the expression above is:

$$L\left(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \rho, \sigma_\varepsilon^2\right) = \sum_i \ln\left(1 - \rho \upsilon_i\right) - \frac{N}{2}\ln\left(2\pi\right) - \frac{N}{2}\ln\left(\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\varepsilon'\varepsilon$$
$$\varepsilon = u - \alpha_1 - \alpha_2 v - \alpha_3 \text{Sect} - \alpha_4 \text{Dem} - \alpha_5 h - \rho Wu$$

where υ_i (i=1,2,...,N) are the eigenvalues of W. Under the usual regularity conditions, the maximisation of the later expression provides with the maximum likelihood estimators of the parameters. To solve the maximisation problem we use the Newton-Rapshon algorithm implemented in the OPTMUM library of the GAUSS system.