THE REGIONAL DISTRIBUTION OF SPANISH UNEMPLOYMENT.

A SPATIAL ANALYSIS.*

E. LOPEZ-BAZO

T. DEL BARRIO

M. ARTIS

Address: Research Group "Anàlisi Quantitativa Regional"

Dept of Econometrics, Statistics and Spanish Economy

University of Barcelona

Avda. Diagonal 690, 08034 Barcelona, Spain

Phone: +34934024320

Fax: +34934021821

Email: elopez@eco.ub.es

* The authors acknowledge financial support by the CICYT, SEC99-0700.

Abstract: This paper proposes a set of tools for analysing the regional distribution of unemployment. As we were interested in the characteristics of the distribution as a whole, results from a traditional regression analysis were complemented with those obtained by estimating its external shape before and after being conditioned to factors underlying regional unemployment. In addition, the paper specifically considers the spatial characteristics of the distribution, and the empirical model developed in order to determine explanatory factors includes spatial effects. This framework is applied to the study of the provincial distribution of unemployment rates in Spain. Results point to increasing spatial dependence in the distribution of regional unemployment rates, and a change in the factors causing regional differentials over the last decade.

Keywords: geography of unemployment, spatial analysis, Spanish regions

JEL codes: C21, E24, R12, R23

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

1. INTRODUCTION

High unemployment rates have typified some European countries in recent decades. Most studies point to structural conditions and rigidities of the labour market, together with the system of unemployment benefit in those European economies, as the major causes of such high figures (Bean, 1994). In addition to the question of nationwide aggregate unemployment, another interesting, but less studied, aspect is the geographical distribution of unemployment. There is, however, evidence as to the relevance of spatial differentials with respect to unemployment rates in Europe, Canada and the US. Aside from the fact that labour markets remain essentially regional, there are reasons for considering unemployment from a regional point of view. Elhorst (2000) proposes three: the magnitude of regional differences between regions within countries; the absence of explanations for the existence of regional unemployment disparities in macroeconomics; and the inefficiency created by such disparities in the economy as a whole. In this regard, most previous contributions have aimed to analyse the determinants of regional unemployment by using a regression analysis, in which unemployment in regions of a given economy is explained by a set of explanatory variables that include characteristics of the regional labour market, of the population, the industrial mix, nationwide unemployment, etc (Marston, 1985; Elhorst, 1995; Partridge and Rickman, 1997; Taylor and Bradley, 1997)[1].

Such analyses provide an estimate of the effect that each factor has on the unemployment rate of *an average or representative region* in the sample being analysed. Quah (1993, 1996) initially raised this point with respect to growth regressions. He suggested studying the effect on the whole distribution of the economic variable under analysis by complementing the

[1] Taylor and Bradley (1997) and Elhorst (2000) discuss the determinants of regional unemployment and

traditional analysis with alternative techniques. This approach has recently been applied to the analysis of the dynamics of regional unemployment rates (Overman and Puga, 2002; López-Bazo et al, 2000). In this paper, we develop it further by combining the results of a regression analysis with the estimation of the shape of the regional distribution of unemployment, conditional to some of the above-mentioned factors. Comparing the real observed distribution with that in which the impact of the explanatory variables has been removed allows their effect on the characteristics of the distribution as a whole to be determined. Our results provide interesting insights into, for example, the formation of groups of regions with separate unemployment rates.

We believe, also, that analyses of regional unemployment should specifically consider the spatial characteristics of the distribution, and empirical models developed in order to determine explanatory factors should include the possibility of spatial effects. Spatial interactions across regional labour markets may be the result of workers in a region being willing to fill vacancies in other regions and firms looking for workers outside the regions in which they are located. Burda and Profit (1996), for local labour markets in the Czech Republic, and Burgess and Profit (2001), for the travel-to-work areas in Britain, have provided evidence for the existence of such spatial interactions. More generally, the outcome of the labour market in a region could be influenced by the circumstances of other regions in the system. In this regard, Bronars and Jansen (1987) and Molho (1995) report the significance of spatial spillovers in the process by which unemployment differentials adjust to local shocks in the UK and the US, respectively.

Accordingly, our study includes an explicit spatial econometric analysis of the regional distribution of unemployment and, therefore, is consistent with the work of Rey and Montoury (1999), who reconsidered the question of regional economic growth from a spatial econometric perspective. Their paper provided new insights into the geographical dynamics

---

summarise results from previous research.

of US regional income growth patterns by applying methods of exploratory spatial data analysis and including spatial effects in the econometric models used to study regional income convergence.

In our paper, we apply the analysis of the regional distribution of unemployment rates, including spatial effects, to Spanish unemployment. Several studies have tried to explain why unemployment in Spain has behaved the way it has, and also, why it has followed a different pattern to that experienced in other countries (Bentolila and Blanchard, 1990; Blanchard and Jimeno, 1995; Dolado and Jimeno, 1997; Marimon and Zilibotti, 1998). However, the regional distribution of unemployment rates in Spain has attracted less attention. Yet, as will be shown below, the Spanish case is somewhat extreme in this regard as well. The distribution of unemployment rates is characterised by sizeable differences between regions and a remarkable stability in their ranking. Thus, the Spanish provinces (NUTS III regions in Spain[2]) with the highest unemployment in the late nineties have rates that are almost double the Spanish average for those years. They were also among the regions with the highest rates in previous decades. In contrast, some other provinces had rates that were actually below the EU average. Indeed, in recent decades they have consistently been among the most favoured provinces in Spain, with rates never above half the Spanish average.

Our analysis is focused on the distribution of unemployment in the 50 Spanish provinces for two particular years, 1985 and 1997. It is interesting to study changes in the distribution over a period in which the Spanish economy underwent important economic reforms as a result of the processes of market liberalisation, openness and integration into the European Union. In addition, labour market reforms in that period were aimed at increasing flexibility and deregulation[3]. It is likely that Spanish regions did not all react in the same way to these

---

[2] Most labour commuting takes place within these territorial units, so they can be taken to approximately define integrated labour markets. The size of the average province, as measured by the labour force, was 571,654 workers in 1985 and 644,879 in 1997.

[3] There have been various reforms in labour market legislation in Spain over the last two decades (1984, 1992,

reforms. In addition, the determinants of unemployment differences across provinces may have changed during that period, and this is in fact confirmed by our results. The first year in our analysis comes at the end of a decade of crisis and industrial restructuring, and was the year before Spain joined the European Community. At that time, unemployment figures reached their highest levels ever. Twelve years later, the Spanish economy had undergone a period of notable growth and a fall in unemployment rates (late eighties and early nineties), followed by some years of deceleration and a rise in unemployment to previous levels. Thus, the two points in time that we are considering encompass a complete cycle and, therefore, the analysis is not contaminated by separate regional responses to the different phases of the business cycle. In addition, similar aggregate unemployment rates for Spain in both years mean that the analyses of relative or absolute deviations do not differ greatly (see Martin, 1997 for a discussion of regional unemployment disparities in terms of *relativities* or *differentials*).

The rest of the paper is organised as follows: a preliminary spatial exploratory analysis of the distribution of unemployment rates in the Spanish provinces is presented in section 2, where the techniques used throughout the paper are concisely described; section 3 briefly summarises the explanatory variables of regional unemployment included in our study and presents the empirical model used in section 4. This section describes the results of the explanatory analysis. It includes the regression results and analysis of the impact of the variables influencing unemployment in the provincial distribution. The paper's final section offers some concluding comments.

1994, 1997). These introduced new types of contracts (part time, training, fixed duration), decreased the cost of firing workers, and redefined the system of unemployment benefit. However, doubts have been raised about the effectiveness of such measures, while none of the reforms addressed the problem of heavily centralised labour market bargaining (see Segura, 2001 for further details).

2. EXPLORATORY ANALYSIS OF THE SPATIAL DISTRIBUTION OF UNEMPLOYMENT RATES

Changes over time in the aggregate Spanish unemployment figures have been widely reported and the causes behind their high level in recent decades have been analysed in previous contributions–(Blanchard and Jimeno,1995; Marimon and Zilibotti,1998). During the sixties, the average rate remained stable at around 2-3%. It climbed moderately throughout the next decade, reaching a figure of around 10% by the beginning of the eighties. The unemployment rate then doubled in a five-year period so that more than 20 out of every 100 workers were unemployed (Table 1). Later, unemployment rates moved in parallel with the business cycle, yet always within a range far above those in other Western economies -it was around 20% in 1997. In this same period, the standard deviation, as a raw measure of unemployment differentials in the Spanish provinces, increased markedly up to the mid-eighties. Afterwards, it remained generally stable[4].

A comparison of unemployment rates in those provinces which each year report extreme values provides a clear picture of the magnitude of the spatial differences. The last row of Table 1 shows the difference in unemployment rates between the provinces with the highest and lowest rates in 1985 and 1997. Using unemployment rates as a rough measure of the probability of being unemployed, these figures reveal that workers in certain provinces were much more likely to be unemployed than those in some other provinces. Furthermore, this probability may be increasing. Indeed, more recent figures seem to indicate that certain provinces in north-east Spain are close to full employment, while at a distance of a few hundred kilometres rates remain above 20%. The provincial distribution of unemployment,

[4] These figures, as well as the ones used throughout the paper on labour market variables, come from the Labour Force Survey (EPA) carried out by the Spanish Statistical Office (INE) following the homogeneous EU-wide methodology of EUROSTAT. The survey defines an unemployed person as someone aged 16 or over who has not been employed that week, but who is available for work and is actively seeking a job. Another major source of unemployment data in Spain is the unemployment records of the National Employment Office (INEM). We have discarded this latter source as only part of the unemployed are registered in the INEM.

however, seems to be characterised by strong, though not perfect, persistence, as the correlation coefficient for unemployment differentials in both periods is 0.79[5].

With the aim of providing further insights into the regional pattern of unemployment rates in Spain, we estimated the density function associated with the distribution of unemployment in 1985 and 1997. This function proxies the shape of the distribution, and actually gives more information than the single measures of position and dispersion do. The density function is estimated non-parametrically by the kernel method. The kernel density estimator replaces the "boxes" in a histogram by smooth "bumps" (Silverman 1986). Smoothing is done by putting less weight on observations that are further from the point being evaluated. More technically, the kernel density estimate of a series X at a point x is estimated by

$$f(x) = \frac{1}{Nh} \sum_{i=1}^{n} \mathrm{K}\left(\frac{x - X_i}{h} \right) \tag{1}$$

where N is the number of observations, h is the bandwidth (or smoothing parameter) and K( ) is a kernel function that integrates to one. The kernel function is a weighting function that determines the shape of the bumps. We have used the Gaussian kernel in our estimates:

$$\frac{1}{\sqrt{2\boldsymbol{p}}} \exp\left(-\frac{1}{2} u^2 \right) \tag{2}$$

where u is the argument of the kernel function. The bandwidth, h, controls the smoothness of the density estimate; the larger the bandwidth, the smoother the estimate. Bandwidth selection is of crucial importance in density estimation, and various methods have been suggested in the literature. In this paper we have used the data-based automatic bandwidth suggested by Silverman (1986, equation 3.31):

$$h = 0.9 N^{-1/5} \min\{s, R/134\} \tag{3}$$

5 The coefficient of a simple regression between unemployment differentials in 1997 and those in 1985 is 0.82, with an R^2 of 63.7%.

where s is the standard deviation and R the interquartile range of the series.

The external shape of two or more distributions can be compared by means of the estimated density functions. More specifically, the change in shape of the distribution over the period under analysis can be assessed by comparing the density function for provincial unemployment rates in 1985 and 1997. However, this method comes up against one of the main drawbacks of this type of analysis, namely, how to test the equality of the distributions from the estimated densities. We have addressed this by applying an overlapping coefficient (OVL). Bradley (1985) and Inman and Bradley (1989) promote the use of OVL as an intuitive measure of substantive similarity between two probability distributions. The closer the OVL is to 1, the more similar the distributions being compared. Confidence intervals can be computed by bootstrap techniques in order to test that samples of unemployment rates in both years were indeed drawn from the same theoretical distribution. Additionally, the OVL can be split into the overlap associated with three ranges of unemployment rates: low, medium and high. Further details of this coefficient are provided in the Appendix.

Figure 1 plots the estimated densities for the difference between the unemployment rate in each province and the average rate for the Spanish economy in the years under analysis. In addition to the high degree of dispersion - already illustrated by the data in Table 1 - the figure would seem to show that the shape of the distribution did not undergo important changes. However, a closer look at both densities reveals a tendency towards the concentration of the mass of probability in particular unemployment rate intervals. The most striking feature is the consolidation of a peak at very high positive differentials in 1997, while another peak may be forming to the left of the distribution. This is confirmed by the OVL which, for the whole range of unemployment rate differentials, has a value of 0.873, below the critical value, and thus the hypothesis that both distributions are similar is rejected. Coefficients for the three intervals indicate that differences in the distribution are due to the range of low (0.804) and, especially, high (0.763) unemployment rate differentials, while similarity in the intermediate interval cannot be rejected.

Summing up, changes in provincial unemployment rates over the period under analysis may have caused the formation of two clusters of provinces. The clearest is the one in the range of unemployment rates far above the Spanish average, while the other, perhaps still being formed, is characterised by low relative unemployment.

The above analysis does not, however, consider the particular spatial location of the provinces. Thus, the impact of geography on the dispersion of the distribution and on the process of cluster formation over the period, detected by means of the estimated densities, cannot be assessed. A similar point has recently been raised in studies dealing with the regional distribution of production, and specific tools have been applied in such cases in order to detect the type and intensity of spatial association (Rey and Montouri, 1999; López-Bazo et al, 1999). The type and intensity of spatial association in the regional distribution of unemployment rates can be easily depicted by an X-Y plot in which the standardised value for each region is represented on one axis and the standardised value in the *neighbouring* regions (spatial lag) on the other – a Moran scatterplot, as suggested in Anselin (1996). In addition, the degree of spatial association can be summarised by means of what is known as Moran's I statistic (Moran, 1948). It is defined as:

$$I = \frac{\sum_{i}^{N}\sum_{j}^{N} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{4}$$

where x_i and x_j are the observations for region i and j of the variable under analysis; $\overline{x}$ is the average of that variable in the sample of regions; and w_{ij} is the i-j element of a row-standardised matrix of weights, **W**. This is an NxN matrix of spatial weights whose characteristic element, w_{ij}, summarises the interaction between regions i and j. Different definitions of interactions cause different **W** matrices. Here, we adopted the simplest, but probably also the most popular, definition: the binary contiguity matrix, whereby the element *i-j* of the weight matrix, w_{ij}=1 --before being row-standardised-- if regions *i* and *j* share a border, and w_{ij} = 0 otherwise[6]. Therefore, the spatial lag is simply the average of the unemployment rate in the neighbouring provinces.

The top panel of Figure 2 shows the Moran scatterplot for unemployment rates in 1985, while the plot for the final year, 1997, is shown at the bottom. Results for the Moran's *I* statistic in each year are also shown in the figure. The position of the provinces in quadrants I and III in the Moran scatterplot corresponding to 1985 indicates that provinces with high unemployment rates have neighbours with the same characteristic, while low-unemployment provinces are more likely to be surrounded by provinces with low values . The positive spatial relationship seems to be even stronger in 1997. Accordingly, the value of Moran's I is significant in both cases and higher for the final year under analysis[7]. Therefore, we can conclude that the regional distribution of unemployment rates in Spain is characterised by intense spatial dependence. Furthermore, it seems to have increased over the last two decades.

In order to shed light on the effect which the observed spatial dependence could have on the characteristics of the distribution detected above, we compared the shape of the distribution of provincial unemployment rates, relative to the average rate in Spain, with that for the difference between the rate in each province and the average rate in the neighbouring provinces, that is, the spatial lag of unemployment rates. If some of the dispersion in the distribution is linked to spatial dependence, then we would expect the latter distribution to be more concentrated. Similarly, if cluster formation is, at least partly, a geographical phenomenon, the distribution of unemployment rates in each province minus the rate in the

[6] It should be stressed that the main results in this paper were not affected by the use of a distance weight matrix. On the contrary, the role of spatial dependence was even larger in that case.
[7] Spatial dependence is observed in each one of the years between 1985 and 1997, with continuous increase. These results are not reported in order to save space; they will be provided upon request.

neighbouring ones should not show the mass of probability at the very high and low unemployment rates. Given that we previously detected some changes in the shape of the distribution, and in the degree of spatial autocorrelation between 1985 and 1997, we made the comparison for both years (Figure 3).

It can be seen how the distribution in 1985 shifts to the right when the neighbouring effect is removed, being the mode located now around zero. It is also moderately more concentrated than the original distribution, although the mass of probability remains at the large positive differentials. The OVL clearly indicates that distributions are different, particularly in the case of the low unemployment range. The same exercise for 1997 reveals that the neighbouring effect could be responsible for most of the characteristics of that year's distribution. Not only is the distribution now less dispersed, but also the clusters detected in the original distribution completely disappear when the density function is estimated for deviations with respect to neighbouring provinces. Once again, conclusions from visual inspection are confirmed by the OVL coefficients.

Summing up, a simple descriptive analysis shows how the regional distribution of unemployment is Spain is largely dispersed, and that there is a trend toward the formation of clusters of extreme values. Furthermore, the spatial distribution is far from random or homogeneous. On the contrary, the unemployment rate in a province is increasingly related to the one in the surrounding provinces, and this phenomenon could be responsible for the majority of the distribution's characteristics. This is particularly so in 1997, where it seems to account for the above-mentioned clusters of provinces.

3. EMPIRICAL MODEL OF REGIONAL UNEMPLOYMENT

Causes of regional unemployment have been discussed in detail in the literature (Marston, 1985; Partridge and Rickman, 1997; Martin, 1997; Taylor and Bradley, 1997). Elhorst (2000) has recently produced a comprehensive review that includes a list of the explanatory variables suggested as having an influence on regional unemployment. Among the factors on the list are the natural change in the labour force, the participation rate, net in-migration and commuting, wages, employment growth, the industrial mix, the educational attainment of the population, market potential, and other characteristics of the labour market such as the degree of unionisation. Although we do not intend to describe in detail the effects which those variables have on regional unemployment rate differentials, we should point out that we used the above-mentioned papers in selecting the variables for our empirical model. The process of selection was also influenced by studies providing particular evidence about factors which affect Spanish unemployment (e.g. Rodríguez-Pose, 1996 and 1998; Marimon and Zilibotti, 1998), and the availability of reliable data at the provincial level. In this regard, we were not able to include factors such as long-term unemployment and unemployment benefits due to the lack of spatially disaggregated data for those variables. However, the omission of these factors should not alter the main results if they have a homogeneous impact on all provinces, given that we are focusing on unemployment differentials. Furthermore, some of the variables already included may capture at least some of their effects. The factors finally selected in our analysis were as follows:[8]

- *Employment growth* (**EMP**): It is expected that additional jobs decrease the unemployment rate, and most of the studies which have considered this variable support that negative effect. However, the sign of the influence can be reversed, as pointed out by Harris and Todaro (1970), through induced urban-rural migration.
- *Net migration* (**M**): The effect of net migration on regional unemployment rates is not straightforward, as it may increase labour supply and demand over a long time period. Accordingly, empirical evidence has produced mixed results. In the case of Spain it

[8] The precise definition of variables and sources can be found in the data appendix.

should be stressed that international as well as interregional migration flows were an important mechanism in balancing the labour market up to the eighties, though they fall to lower levels in the last two decades.

- *Unit labour costs* (**ULC**): We assumed that firms are concerned with wages in relation to labour productivity, since wage differences across regions accommodate to productivity differences. In so doing, we are basically considering the positive influence of labour costs on unemployment through the effect on labour demand. The impact through labour supply would require the use of data on real wages. As far as we know, series on provincial prices are not available to correct nominal labour costs for differences in purchasing power. Nominal labour costs were always non-significant when they were introduced in the analysis.
- *Industrial mix*: We controlled for the share of agriculture (**%AGR**) and manufacturing (**%MANU**) in employment. Regions specialised in declining industries are expected to exhibit higher unemployment rates than those based around growing activities. Industrial restructuring in the seventies and eighties was particularly severe in Spain. As a result, employment in agriculture and manufacturing fell markedly. Consequently, a negative relationship between employment share in those industries and unemployment rates would be expected. However, Elhorst (2000), considering the possibility of industrial mismatch and some drawbacks in the use of employment shares, points out that it is not clear what the sign of these variables should be. This is confirmed by the diversity of results obtained by empirical studies which have included these variables (e.g. Elhorst, 1995; Partridge and Rickman, 1995 and 1997;Taylor and Bradley, 1997).
- *Human capital* (**H**): For a number of reasons (higher demand for skills, lower probability of lay off, influence on migration decisions, etc) the educational attainment of workers is expected to be negatively related to unemployment rates. Unemployment rates for

workers with higher level studies have been reported to be lower than for workers who leave education with few or no qualifications (Nickell and Bell, 1996). There has been a constant increase in the level of education of the Spanish population over recent decades, but regional differences in these levels remain great (Rodríguez-Pose, 1996). If the average human capital of the labour force in the Spanish provinces differs, this might explain some of the inequality in the geographical distribution of unemployment. We have proxied this factor by the percentage of the labour force that has at least started secondary schooling.

- *Demography and participation*: The structure of the population has an obvious effect on labour supply. Unemployment rates have been notably higher for people aged 16-25. In the Spanish economy, 36 out of 100 workers aged under 25 were unemployed in 1999 - well above the 19% average for the EU as a whole. Furthermore, differences across Spanish regions are notable: above 40% in those with higher youth unemployment and below 25% where the problem is less intense. Therefore, our model includes the share of working age population aged 16 to 25 (**YOU**). As regards participation, there is a controversy about the effect of participation rates on unemployment, as several opposite mechanisms might be at work simultaneously (Elhorts, 2000). To allow for the possibility that these mechanisms exerted a separate influence on male and female participation decisions, we included both participation rates (**MALE**, **FEMALE**) as explanatory variables in our model.

As a result the model to be estimated can be expressed as:

$$\mathbf{U}^t = \beta_0 + \beta_1\,\mathbf{EMP}^t + \beta_2\,\mathbf{M}^t + \beta_3\,\mathbf{ULC}^t + \beta_4\,\mathbf{\%AGR}^t + \beta_5\,\mathbf{\%MANU}^t + \beta_6\,\mathbf{H}^t + \beta_7\,\mathbf{YOU}^t + \beta_8\,\mathbf{MALE}^t + \beta_9\,\mathbf{FEMALE}^t + \boldsymbol{\varepsilon}^t \qquad (5)$$

where $\mathbf{U}^t$ is the vector of differences between the unemployment rate in each province in year t (=1985,1997) and the average unemployment rate in Spain. The explanatory variables, as defined above, are all expressed as deviations from the Spanish average as well. Finally, ε is a random perturbance. The unknown coefficients were estimated by ordinary least squares (OLS) using the observations from the 50 Spanish provinces. Given that the effect on unemployment rates of the explanatory variable may have changed over the period under analysis, we did not impose equality restrictions on the coefficients across equations for each one of the years.

However, we did check for spatial dependence in the residuals of the regressions for each one of the years. Three tests of spatial dependence were computed: the residuals Moran's I, and the robust Lagrange multiplier tests for spatial lag and spatial error autocorrelation. While the Moran test is not able to distinguish the two types of spatial autocorrelation, the robust tests have been shown to have good power against a specific alternative (Anselin et al, 1996), and thus can be used to formulate the appropriate spatial model (Florax and Folmer, 1992). More specifically, the spatial error model considers the following structure for the perturbance of (5):

$$\mathbf{e} = d\mathbf{W}\mathbf{e} + \mathbf{x} \tag{6}$$

where $\mathbf{e}$ is the perturbance vector, $\mathbf{W}$ the matrix of spatial weights defined in the previous section, d the spatial error coefficient, and $\mathbf{x} \sim \mathrm{N}(\mathbf{0}, s^2\mathbf{I})$. The spatial lag model includes the spatial lag of the unemployment rates (**WU**) in the list of regressors:

$$\mathbf{U}^t = b_0 + b_1\,\mathbf{EMP}^t + b_2\,\mathbf{M}^t + b_3\,\mathbf{ULC}^t + b_4\,\mathbf{\%AGR}^t + b_5\,\mathbf{\%MANU}^t + \tag{7}$$

$$b_6\,\mathbf{H}^t + b_7\,\mathbf{YOU}^t + b_8\,\mathbf{MALE}^t + b_9\,\mathbf{FEMALE}^t + g\,\mathbf{WU}^t + \mathbf{e}^t$$

where g is the spatial autoregressive parameter.

4. EXPLANATORY ANALYSIS OF THE SPATIAL DISTRIBUTION OF UNEMPLOYMENT RATES

In this section we study the influence which the factors outlined in the previous section have on the main characteristics of the distribution of regional unemployment differentials in the two years under analysis, and particularly on the dispersion and clustering described in the exploratory analysis. Several variables proposed in the literature as affecting the level of regional unemployment are considered. These are factors within each province that may influence the performance of the labour market in general, and the rate of unemployment in particular. Given that we have already shown spatial dependence to be an important characteristic of the provincial distribution of unemployment, we also consider the likely existence of interactions across provinces which may help in understanding unemployment rates.

As our interest was not only focused on a representative or average province, we estimated the effect of those factors on the whole distribution of unemployment rates. Therefore, we began with a traditional regression analysis in which estimates of the parameters should provide evidence about the effect which the different variables have on the unemployment differentials of an *average Spanish province* for each of the two years being analysed[9]. Then, using the tools described in section 2, we complemented that analysis by comparing the original distribution with that conditional to the factors under analysis. In so doing, we were able to assess their impact on the whole range of unemployment rate differentials, including, for example, their contribution to the formation of clusters in the distribution.

4.1. Regression results.

[9] Pooling observations for both years would allow unobservable regional effects in unemployment differentials to be accounted for. However, this would be at the cost of imposing equality constraints on the effects of the variables under analysis across time. This hypothesis was clearly rejected by standard tests.

We applied OLS to the linear specification given by (5), although the dependent variable was restricted to the interval {-u_{NAT}, 1-u_{NAT}}, where u_{NAT} is the nationwide unemployment rate[10]. This is a common problem in empirical analyses of unemployment rates, and only a few studies have applied the logistic transformation in order to address this (see the summary of the collection of studies in Table 1 in Elhorst, 2000). When the focus of the analysis is the regional unemployment rate, the dependent variable ranges within the interval {0,1}, and can be taken to be the probability of an average worker in a region being unemployed. Thus, the model proposed for analysing regional unemployment rates is based on proportions data, and so the logistic transformation is appropriate. Unfortunately, in our case, such a transformation could not be applied as regional unemployment rate differentials may be negative. Therefore, we continued to estimate the coefficients based on the linear model, but reported the standard errors from the White (1980) heteroskedasticity consistent estimator of the covariance matrix for the parameter estimates. In so doing, we sought to account for the heteroskedastic perturbance of a model of proportions data (see Greene, 1993 for further details).

The OLS estimates of (5) for 1985 and 1997 are summarised in Table 2. Before discussing the sign and magnitude of the estimated coefficients, it should be stressed that the overall fit achieved by the factors included in the specification for both years is quite high. Furthermore, the degree of collinearity among the regressors, as summarised by the condition number, is surprisingly moderate, taking into account the cross influence of the different factors. This enables us to be more confident in the estimates of single coefficients. However, the spatial dependence tests point to the presence of spatial autocorrelation in the residuals of the equations for both years. In accordance with the results in the exploratory section, this phenomenon seems to be more intense in 1997. As spatial autocorrelation would invalidate conclusions based on the misspecified model, we have not commented on the value of the parameters from the OLS estimates. Instead, we have estimated the model which best

[10] We thank an anonymous referee for pointing this out to us.

accounts for spatial dependence. In this regard, the values of the robust tests clearly point to the spatial lag model as the preferred specification. However, the OLS is inconsistent in this case due to simultaneity induced by the spatial lag (Anselin, 1988). Instrumental variables and maximum likelihood estimators have been suggested to provide consistent estimates.

Table 3 presents the maximum likelihood estimates of the spatial lag model (7), where reported standard errors come from the heteroskedasticity consistent estimator of the covariance matrix of the maximum likelihood parameters, as suggested in White (1982) and Davidson and MacKinnon (1993). The major conclusion to be drawn from the parameters is a change in the main causes of provincial unemployment rate differences. While excess of labour costs over productivity, industrial mix, and human capital differences across provinces seem to explain most of the provincial unemployment rates in the mid-eighties, they lose their explanatory power at the end of the nineties. Unit labour costs affect positively, and human capital negatively, the rate of unemployment in 1985, as expected on a priori grounds. Differences in the share of manufacturing employment, and particularly of agriculture, have significant coefficients in 1985. They show a negative effect on unemployment differentials which, despite being somewhat counterintuitive, is in line with results obtained for some other economies (Jones and Manning, 1992; Taylor and Bradley, 1997; and the discussion in Elhorst, 2000). However, neither the change in employment nor the demography and participation variables have a significant impact on unemployment differentials for that year.

In sharp contrast, the variables with significant coefficients at the usual levels in 1997 are employment growth, net migration, youth population and female participation. Provinces that create employment at higher rates tend to experience less relative unemployment. The same applies to net in-migration, as supply-side effects seem to surpass the demand-side effects and, therefore, provinces with a net increase in people had, conditional to the other factors, lower unemployment rates in 1997. The positive effect which the percentage of youth population has on unemployment differentials is particularly strong. One extra point of difference between a province and the national average translated into more than one additional point in the difference in unemployment rates. Finally, although the effect of male participation rates is negligible, female participation reduces unemployment rates. This could be due to the fact that female decisions to participate in the Spanish labour market are closely related to the current level of unemployment. As a result, female participation would be lower in provinces with high unemployment and higher where unemployment is low. In any case, there is an important dispersion in this estimated effect as the coefficient is only significant at 10%.

Thus, none of the factors that appear to be significant in explaining unemployment rate differentials for an average Spanish province in the mid-eighties seems to be important in the late nineties. There is another noteworthy result from these estimates, namely, the increase in the spatial coefficient observed over the period. Its value is estimated to be 0.284 and significant at 5% by a t-ratio test in 1985. However, it is only significant at 10% when a more appropriate likelihood ratio test (LR-LAG) is used. Therefore, we can conclude that most of the spatial dependence detected in the provincial distribution of unemployment rates can be explained by factors within each of the provinces included in our empirical model. On the contrary, the spatial coefficient in 1997 is clearly significant, and is double the one for the initial year. Furthermore, there is no evidence of remaining spatial autocorrelation in the residuals (LM-ERR).

As a matter of comparison, Table 3 includes the estimates of a pure autoregressive spatial model - excluding from our specification the factors within each province. In this case, the spatial parameter for both years is quite similar and as high as 0.669 and 0.751, respectively, in accordance with the exploratory results above. The estimated value for the spatial coefficient is, therefore, much lower when factors within each province are included, indicating that spatial dependence in the explanatory variables was mostly responsible for spatial dependence in the distribution of unemployment rates in 1985, although only partially responsible for this phenomenon in 1997.

4.2. Conditioned distributions.

Once an estimate of the parameters in (7) was available we could obtain the distribution of relative unemployment conditional to the factors defined above for each of the years. In order to separate the effect of within-province factors from the spatial effect, we computed a conditional distribution for each. To do this, we first had to compute the unemployment differentials conditional to the set of factors. This was obtained by combining the estimates for the parameters and the corresponding variables plus the vector of residuals, where the values for the variables we wanted to condition for were set to zero. That is to say, we estimated the unemployment differentials in case there was no difference across provinces with respect to the factors within each region that affect unemployment, leaving unaltered the original values for the spatial lag of unemployment rates. Correspondingly, the distribution conditioned to having *similar neighbours* was obtained by substituting the values of the spatial lag for a vector of zeros, while using current values for the other variables in the model. The density function for the unconditional and conditional distributions could then be computed as described above. Visual inspection of both densities for each year and the calculation of the OVL coefficients enabled their similarity to be checked and conclusions could thus be drawn about the impact of the variables on the whole distribution.

Figure 4 depicts the densities for the current distribution of unemployment differentials and the distribution conditioned to no differences in within-province factors, whereas those for the distribution conditioned to the neighbouring effect are shown in Figure 5. In the first year (top panel), it can be observed how the geographical distribution of unemployment would have been much more concentrated had the provinces not differed in the rate of employment growth, migration flows, unit labour costs, industrial mix, educational attainment, youth population and participation rates. In fact, the conditional distribution almost collapses around the range of no differences. As a result, the OVL coefficient rejects similarity between the unconditional and conditional distributions, for the whole range and for the three intervals defined above. Moreover, factors within each region almost completely explain the mass of probability at the positive differentials detected in the real distribution. In contrast, there are no significant differences between the real distribution and the one that results from removing the spatial lag effects. The only noteworthy effect is observed in the interval of positive differentials. When the neighbouring effect is removed the mass of probability in that interval shifts to the left. In fact, the OVL_{HIGH} leads to rejection of similarity for that particular interval, and is strong enough to cause the global OVL to reject similarity for the whole range, even when similarity seems to be acceptable for low and medium unemployment differentials.

The picture for the end of the nineties (bottom panel of Figures 4 and 5) shows, once again, how factors within each province account for an important amount of the distribution's characteristics. Once conditioned, most of the probability is concentrated close to the point of no regional differences. The cluster of low relative unemployment disappears and the one of positive differentials shifts to the left. However, it is also clear that these factors cannot fully explain the cluster. Interestingly, it is mostly explained by the spatial interaction effect, as shown by the distribution once conditioned to no differences across provinces in the spatial lag of unemployment.

Summing up, the within-province factors considered in our study account for most of the distribution's characteristics in 1985, the neighbouring effect having only a moderate influence. This latter effect only helps to explain some aspects of the cluster of positive differentials under formation that year. Of greater importance, however, is its explanatory value with respect to such phenomena in 1997, as here the cluster is almost unexplained by the explanatory variables. Nevertheless, they still seem to be responsible for most of the dispersion in the distribution.

5. CONCLUSION

This paper has analysed the distribution of unemployment in the Spanish provinces from a new perspective, and has paid special attention to the spatial dimension of the phenomenon. A set of statistical tools for studying both changes over time in the distribution of unemployment rates and the influence which the determinants of regional unemployment have on the whole distribution has been proposed. Furthermore, spatial effects in that distribution have been specifically analysed by applying spatial exploratory techniques and spatial econometric models. Our results for the Spanish provinces show how this type of study complements the traditional regression analysis and provides new insights into the geographical distribution of unemployment.

Applying the above to the Spanish provinces for the last two decades reveals that the ongoing processes of economic integration and labour market deregulation have caused a kind of regional cluster formation, as the distribution of unemployment rates in the late nineties shows a mass of probability at the interval of large relative unemployment, while another group of regions, where rates are far behind the nationwide average, may also be under formation. This would confirm differences in the regional reaction to the new economic framework. Interestingly, our results reveal a shift in those factors which may explain unemployment differentials from the mid-eighties to the late nineties. While, in 1985, the significant variables were differences in unit labour costs, the industrial mix and, to a lesser degree, the educational attainment of the labour force, these do not explain the main characteristics of the distribution in 1997. In contrast, the dispersion in the distribution for that year seems to be related to a region's ability to create employment, to the net attraction of population, and characteristics of the regional population, such as the percentage of youth population or female participation in the labour market. It should be stressed, however, that contrary to what happened in 1985, spatial effects play a role at the end of the nineties. Aside from the fact that the spatial lag of unemployment rates could well be proxying for other factors within each region that were not included in our analysis, spatial effects prove to be highly significant in the regression analysis and almost completely account for the cluster of provinces with unemployment rates above the average. Therefore, it would be worthwhile including explicitly spatial variables in future empirical analyses of regional unemployment in order to elucidate which kinds of mechanism are responsible for the significant spatial effects detected in this paper.

Finally, we would like to stress that other economies in Europe may share at least some of the characteristics observed in the case of Spain. Large dispersion in regional unemployment rates not only characterises the European Union as a whole, but is also common to some member states. Policies aimed at alleviating this problem can only be developed if the reasons for such spatial disparities in these economies are clearly understood. We therefore aim to carry out further studies in the future.

APPENDIX

A.1. Data description and sources

Variable	**Definition**	**Source**
Unemployment rate	$U_i = \frac{u_i}{A_i} * 100$ u_i: Total unemployed force in region *i* A_i: Total labour force in region *i*	Labour Force Survey (EPA) from the Spanish Statistical Office (INE) http://www.ine.es
Employment growth	$EMP_i = \frac{L_{it} - L_{it-5}}{L_{it-5}}$ L_{it}: Employment in region *i* in period *t*	EPA
Migration	$M_i = \frac{NM_i - OM_i}{POP_i}$ NM_i: in-migration OM_i: Out-migration POP_i: Population	INE
Labour unit costs	$UCL_i = \frac{LC_i}{GDPPW_i}$ LC_i: Labour costs per worker $GDPPW_i$: Gross domestic product per worker	Fundación BBVA http://bancoreg.fbbva.es.
Labour in agriculture	$\% AGR_i = \frac{agri_i}{L_i} * 100$ $agri_i$: Employment in agriculture in region *i* L_i: Total employment in region *i*	EPA
Labour in manufacturing	$\% MANU_i = \frac{manu_i}{L_i} * 100$ $manu_i$: Employment in manufacturing in region *i* L_i: Total employment in region *i*	EPA
Human capital	$H_i = \frac{h_i}{A_i} * 100$ h_i: Labour force that has at least started secondary schooling in region *i* A_i: Total labour force in region *i*	From Pérez and Serrano (1998), taking as primary source EPA
Female participation	$FEMALE_i = \frac{FEML_i}{FEM16-65_i} * 100$ $FEML_i$: Female labour force in region *i* $FEM16-65_i$: Females of working age in region *i*	EPA
Male participation	$MALE_i = \frac{MAL_i}{MAL16-65_i} * 100$ MAL_i: Male labour force $MAL16\text{-}65_i$: Males of working age	EPA
Youth population	$YOU_i = \frac{you_i}{N16-65_i} * 100$ you_i: Population aged 16 to 25 years old in region *i* $N16-65_i$: Population of working age in region *i*	EPA

A.2. 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 used the OVL to compare frequencies throughout the range of a variable for two samples. The idea behind the OVL can be summarised in the following figure, where the range of values of two variables, x1 and x2, is on the x-axis, and the density on the y-axis. The OVL is the area where the densities of the two distributions overlap when they are plotted on the same axes.

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

$$OVL = \sum_{x} \min\left[\mathrm{f}(x_1), \mathrm{f}(x_2)\right] \qquad \text{(A.1)}$$
$$OVL \in [0,1]$$

where $\mathrm{f}(x_1)$ and $\mathrm{f}(x_2)$ are the empirical density functions. In the case of continuous distributions, summation is replaced by integration. A value of 1 for the OVL means that the two density functions are exactly the same, whereas a null value indicates the absence of overlapping in the density function at any point in the range of the variable. The closer the OVL to 1 the more similar the two distributions being compared.

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:

$$OVL_a = \frac{\sum_{x \in a} \min\left[\mathrm{f}(x_1), \mathrm{f}(x_2)\right]}{\sum_{x \in a} \max\left[\mathrm{f}(x_1), \mathrm{f}(x_2)\right]} \in [0,1] \qquad \text{(A.2)}$$

where a denotes a specific interval.

We have computed the OVL_a for three different intervals of the unemployment rate differentials (a=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 coefficient depend on those of the data under analysis. Thus, the way to approach the issue is via simulation. Furthermore, the OVL is a biased statistic, because any sampling variation in the densities of two samples obtained from the same population causes the OVL to be strictly less than one.

We used the bootstrap method to obtain the mean and variance of the OVL. We did this by resampling both the original data and a simulated sample of the same size from a

$N(\bar{x}_i, s_i)$, (i=85, 97). The number of replications is m=10000. Tabulated values in Tables A.1 and A.2 were used to construct a kind of confidence interval in order to test the hypothesis of equality of two distributions. The rule of thumb was to reject the hypothesis of similar distributions if the value estimated for the *OVL* was lower than the expected value for the *OVL* in each case minus twice the standard deviation. The null hypothesis was rejected when the overlap was lower than that which would be expected by allowing for sample deviations given the size of our sample. On the contrary, when the *OVL* was closer to 1 than the critical value we could be more confident about assuming similarity.

REFERENCES

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

Anselin L (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fisher M, Scholten H, Unwin D (eds) *Spatial Analytical Perspectives on GIS*. Taylor and Francis, London.

Anselin L, Bera A, Florax, RJGM, Yoon M (1996) Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26:77-104.

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

Bentolila S, Blanchard O (1990) Spanish Unemployment. *Economic Policy* 10: 233-81.

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

Bradley EL Jr (1985) Overlapping coefficient. In Kotz S, Johnson NL (eds.) *Encyclopedia of Statistical Sciences*, (6), 546-547.

Bronars SG, Jansen DW (1987) The geographical distribution of unemployment rates in the US. A spatial-time series analysis. *Journal of Econometrics* 36:251-279.

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

Burgess S, Profit S (2001) Externalities in the matching of workers and firms in Britain. *Labour Economics* 8:313-333.

Davidson R, MacKinnon JG (1993) *Estimation and Inference in Econometrics*. Oxford University Press, Oxford.

Dolado JJ, Jimeno JF (1997) The Causes of Spanish Unemployment: A Structural VAR Approach. *European Economic Review* 41: 1281-1307.

Elhorst JP (1995) Unemployment disparities between regions in the European Union. In: Armstrong HW, Vikerman RW (eds) *Convergence and Divergence among European Regions*. Pion, London.

Elhorst JP (2000) The mystery of regional unemployment differentials: a survey of theoretical and empirical explanations. Research Report Nº 00C06, SOM, University of Groningen, The Netherlands.

Florax R, Folmer H (1992) Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators. *Regional Science and Urban Economics* 22:404-432.

Greene WH (1997) *Econometric Analysis*. Prentice Hall, Upper Saddle River, New Jersey.

Harris JR, Todaro MP (1970) Migration, unemployment and development: a two-sector analysis. *American Economic Review* 60:126-142.

Inman HF, Bradley EL Jr. (1989) The overlapping coefficient as a measure of agreement between two probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics – Theory and Methodology* 18:3852-3874.

Jones DR, Manning DH (1992) Long term unemployment, hysteresis and the unemployment–vacancy relationship: a regional analysis. *Regional Studies* 26:17-29.

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

López-Bazo E, del Barrio T, Artís M (2000) The geographical distribution of unemployment. Paper presented at the XXV European Meeting of the Regional Science Association, Barcelona, August.

Marimon R, 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.

Marston ST (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.

Moran P (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society B* 10:243-251.

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

Overman HG, Puga D (2002) Unemployment clusters across European regions and countries. *Economic Policy*, forthcoming.

Partridge MD, Rickman DS (1995) Differences in state unemployment rates: the role of labor and product market structural shifts. *Southern Economic Journal* 62:89-106.

Partridge MD, Rickman DS (1997) The dispersion in US unemployment rates: The role of market and nonmarket equilibrium factors. *Regional Studies* 31:593-606.

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

Quah D (1993) Empirical cross-section dynamics in economic growth. *European Economic Review* 37:426-434.

Quah D (1996) Convergence empirics across economies with (some) capital mobility. *Journal of Economic Growth* 1:95-124.

Rey S, Montouri BD (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.

Rodríguez-Pose A (1998) *The Dynamics of Regional Growth in Europe: Social and Political Factors*. Oxford University Press, Oxford.

Segura J (2001) La reforma del mercado de trabajo español: un panorama. *Revista de Economía Aplicada* 9: 157-190.

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

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

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48:817-838.

White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50:1:26.

Table 1. Evolution of unemployment rates in the Spanish provinces.

	1985	**1997**
Nationwide unemployment rate	21.63	20.80
Standard deviation	6.80	6.99
Difference between maximum and minimum rates	26.98	29.85

Notes: maximum and minimum correspond to unemployment rates in the provinces with the highest and lowest rates for each year. Thus, they do not necessarily correspond to the same province each year.

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

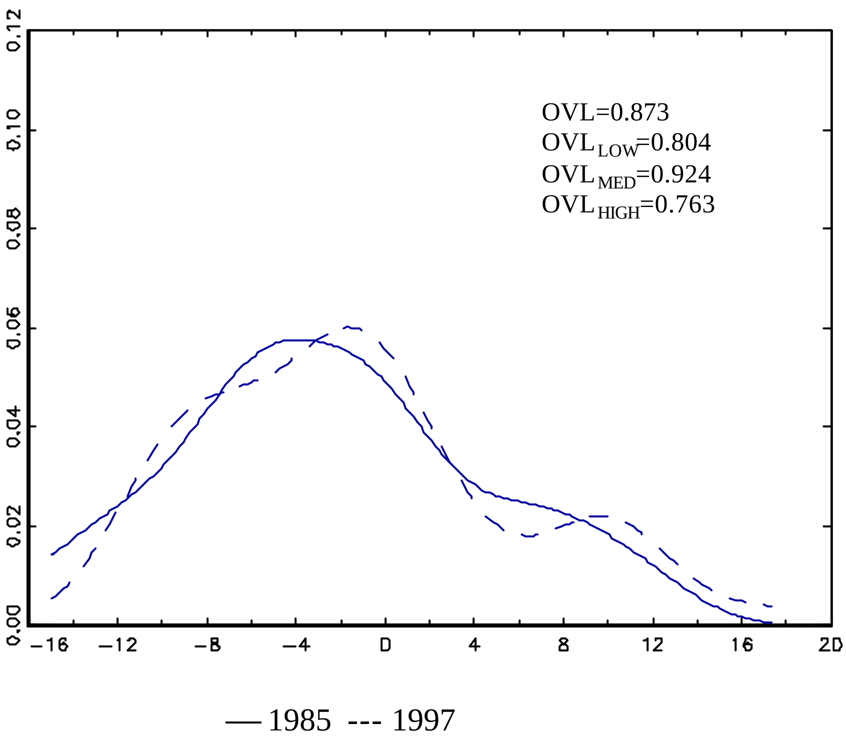
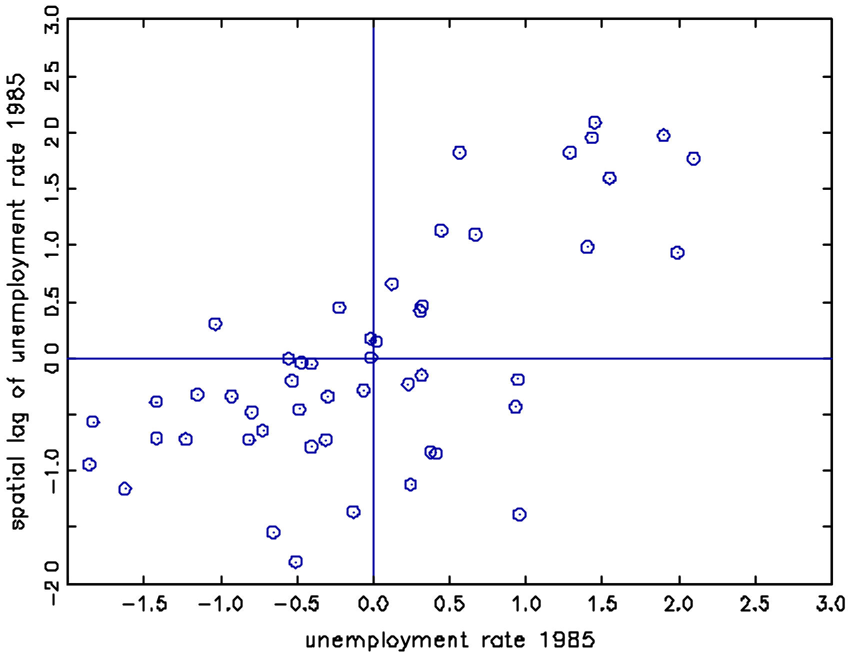


Figure 2. Moran scatterplots of regional unemployment rates.

1985 I-Moran: z-value=6.315 p-value=0.000



1997 I-Moran: z-value=7.298 p-value=0.000

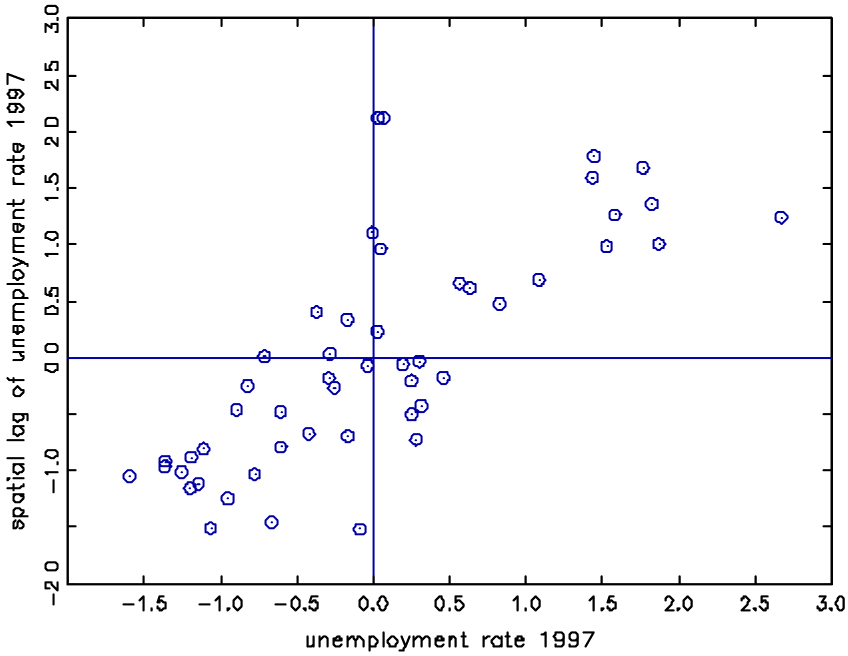
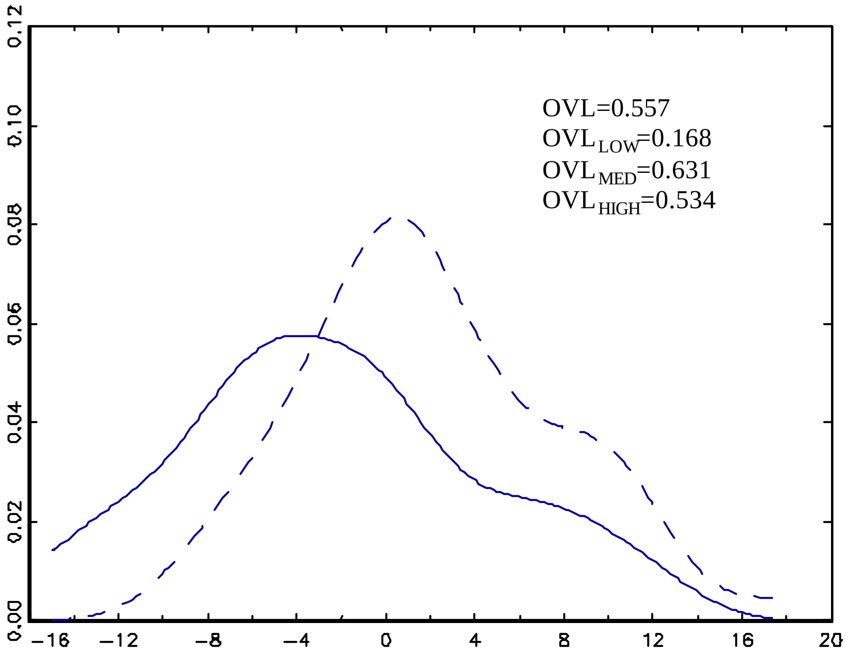
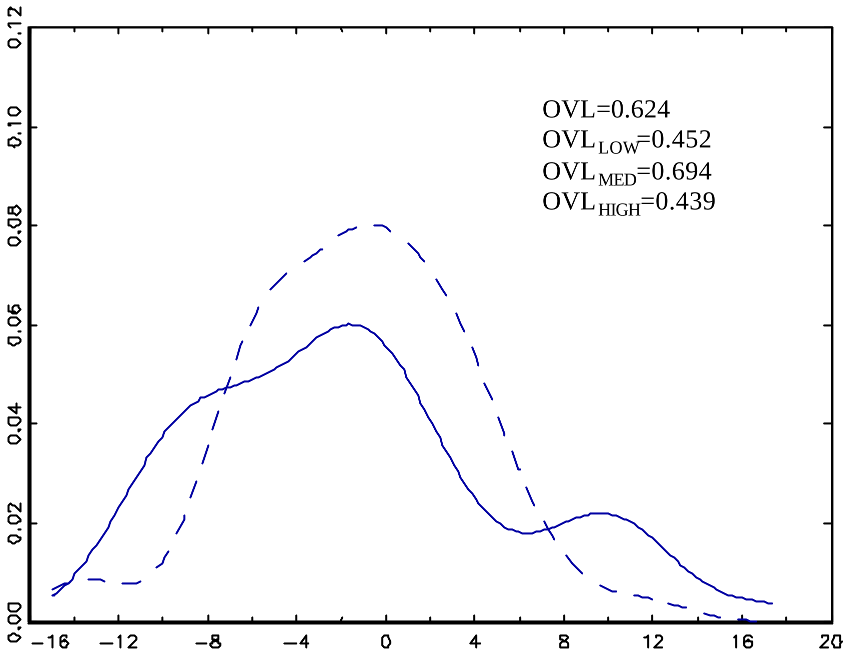


Figure 3. Neighbouring effect in the regional distribution of unemployment rates.

1985



1997



—— aggregate deviation --- neighbourhood deviation

Table 2. OLS estimates of the determinants of regional unemployment and spatial tests.

	1985	1997
Constant	-0.367	0.098
	(0.586)	(0.851)
Employment change	0.226*	-0.221
	(0.135)	(0.162)
Net migration	2.866	-5.829**
	(2.271)	(2.357)
Labour costs	42.719***	17.504
	(12.720)	(17.455)
Share agriculture	-0.425***	-0.060
	(0.125)	(0.119)
Share manufacturing	-0.240*	-0.133
	(0.127)	(0.126)
Human capital	-0.331**	-0.156
	(0.148)	(0.152)
Youth population	0.633	1.693**
	(0.460)	(0.755)
Female participation	-0.112	-0.379*
	(0.145)	(0.204)
Male participation	-0.187	0.110
	(0.288)	(0.411)
R^2	0.749	0.700
LIK	-131.771	-137.585
AIC	283.542	295.171
Condition number	6.624	5.578
Moran's I	0.581	1.944**
Robust LM-ERR	4.042**	3.683*
Robust LM-LAG	6.970***	12.312***

Note: Robust standard errors (White, 1980) in parenthesis. ***, ** and * denote significant at 1%, 5% and 10%, respectively.

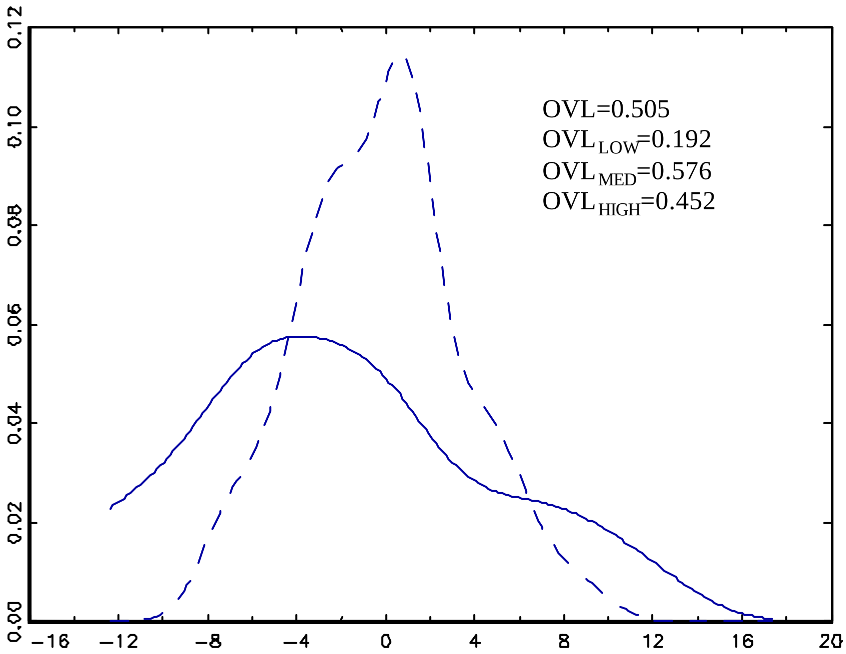
Table 3. ML estimates of the determinants of regional unemployment.

	1985		1997	
Unemployment spatial lag	0.669*** (0.105)	0.284** (0.133)	0.751*** (0.102)	0.462*** (0.141)
Constant	-0.701 (0.656)	0.480 (0.559)	-0.188 (0.660)	0.648 (0.629)
Employment change		0.187 (0.115)		-0.308** (0.136)
Net migration		2.418 (1.980)		-5.106** (2.293)
Labour costs		46.110*** (11.846)		10.752 (10.624)
Share agriculture		-0.346*** (0.100)		-0.054 (0.088)
Share manufacturing		-0.192* (0.113)		0.001 (0.085)
Human capital		-0.223** (0.109)		-0.066 (0.109)
Youth population		0.386 (0.319)		1.283*** (0.456)
Female participation		-0.041 (0.116)		-0.264* (0.161)
Male participation		-0.293 (0.235)		-0.001 (0.295)
LIK	-155.347	-130.056	-151.933	-132.747
AIC	314.693	282.112	307.865	287.494
LR-LAG	22.013***	3.430*	31.533***	9.676***
LM-ERR	1.664	2.854*	4.903**	2.609

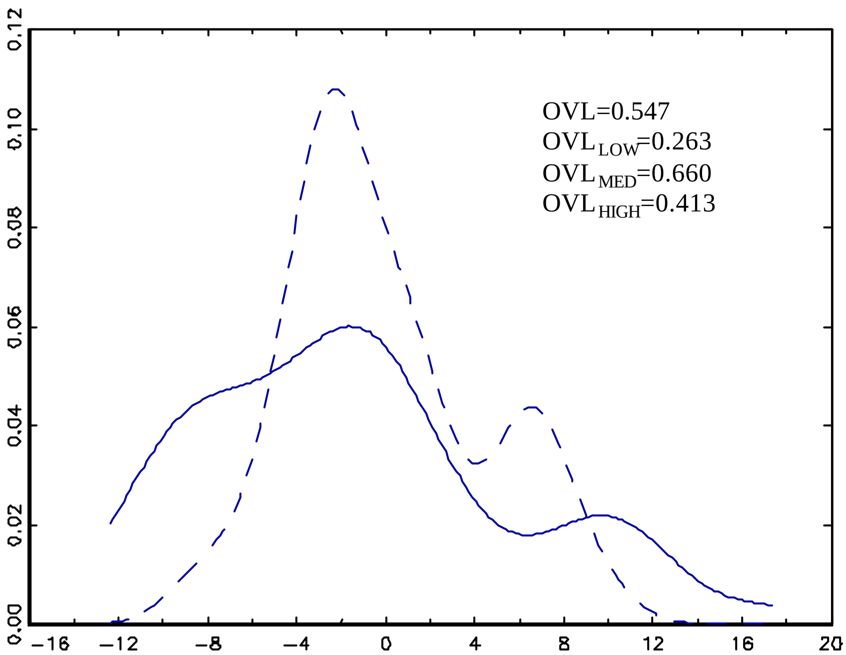
Note: Robust standard errors (White,1982) in parenthesis. ***, ** and * denote significant at 1%, 5% and 10%, respectively.

Figure 4. Distributions conditioned to the determinants of regional unemployment rates.

1985



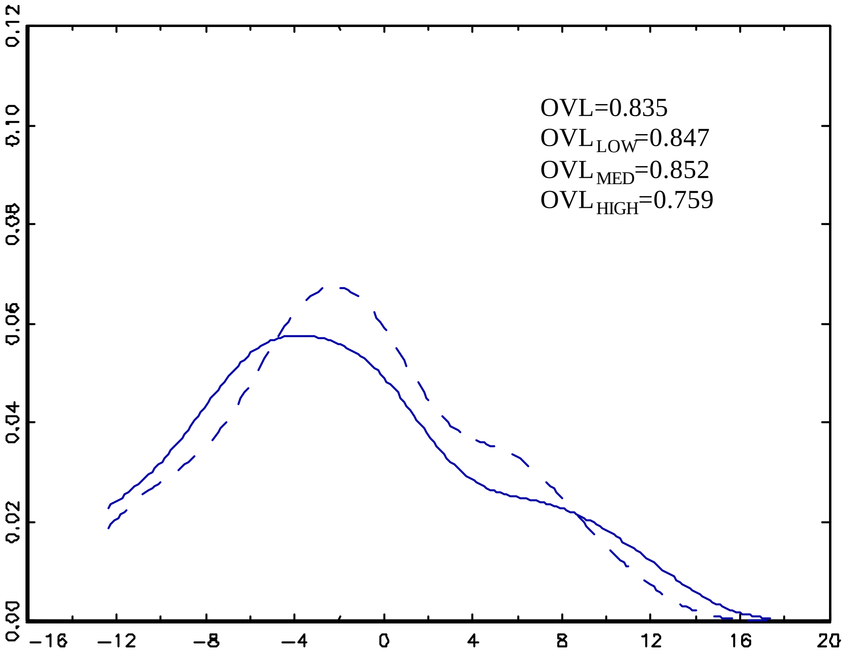
1997



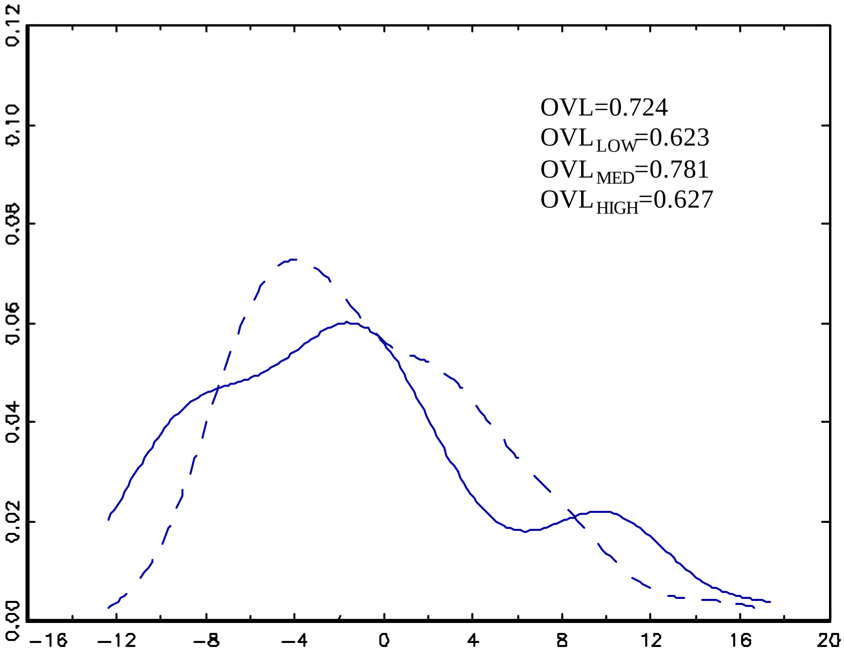
—— real distribution --- conditioned distribution

Figure 5. Distributions conditioned to the neighbouring effect.

1985



1997



— real distribution --- conditioned distribution

Figure A.1. Overlapping Coefficient.

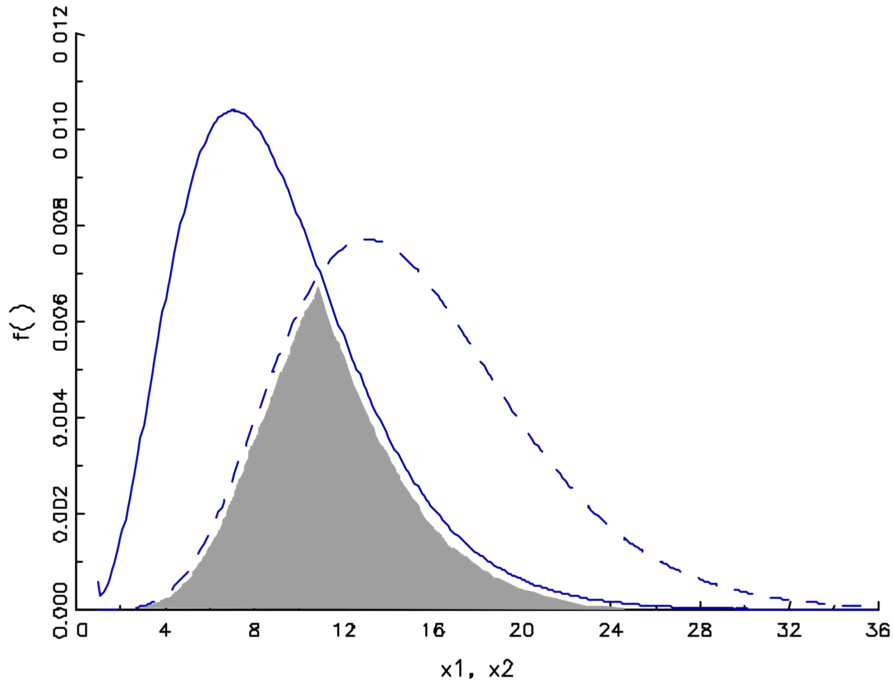


Table A.1. Expected value and variance of the *OVL* by bootstrapping (10000 replications)

	Resample over original data		Resample over $\mathrm{N}(\bar{x}_i, s_i)$ 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 the OVL_a 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