FORECASTING REGIONAL TIME SERIES: THE CASE OF EMPLOYMENT IN EUROPEAN REGIONS

Ana Angulo (*)

Jesús Mur (*)

Fernando López (**)

(*) Department of Economic Analysis.
University of Zaragoza
Gran Vía, 2-4. (50005). Zaragoza. Spain.
e-mails: aangulo@unizar.es, jmur@unizar.es

(**) Department of Quantitative Methods and Computing
Technical University of Cartagena.
Paseo Alfonso XIII, 50 - 30203 Cartagena. Spain.
e-mail: fernando.lopez@upct.es

Abstract

Mur et al. (2008) show that it is relatively simple to obtain symptoms of instability from a model with problems of heterogeneity in cross-sectional spatial model. In this sense, Lopez et al. (2009) propose the use of the local estimation for detecting such situation of instabilities. In the line of previous works, in this paper we try to analyse the capacity of spatial panel data models to deal with the problem of heterogeneity in spatial data. Furthermore, we try to asset whether or not the local estimation technique can be of help for the case of panel data models. We pay special attention to the forecast performance of several alternative models. The empirical application refers to the explanation of employment in European Regional at NUTS II administrative level in terms of Eurostat. Panel data models are estimated on the basis of annual data (1980-2004) and data for 2005-2008 are gathered for evaluating the forecasting performance of the alternative models. The obtained results show that although panel data models are indeed designed for capturing the unobservable heterogeneity of data, the local estimation technique can also be of great help in the context of panel data models. From a forecasting point of view, the best model is the dynamic fixed effect with a spatial lag structure in the equation estimated through local estimation techniques.

Keywords: Dynamic spatial panel data models, Local estimation, European employment, forecast.

JEL Classification: C21; C22; C23; C53; R15

FORECASTING REGIONAL IME SERIES: THE CASE OF EUROPEAN EMPLOYEMENT

1. Introduction

The field of panel data models has received considerable attention during the last decade. Static panel data literature offers the opportunity of allowing for unobservable cross-sectional and time-period specific effects. In addition, dynamic panel data models also offer the possibility of considering the serial dependence between observations on each cross-sectional unit over time. Other advantages of panel data are that they are generally more informative and contain more variation and less collinearity between variables. The use of panel data leads to a greater availability of degrees of freedom and, hence, increases the efficiency of the estimation. Panel data also allow for the specification of more complicated behavioural hypotheses, including effects that cannot be addressed using pure cross-sectional or time-series data (Wooldridge, 2002; Arellano, 2003; Hsiao, 2003; Baltagi, 2005).

When cross-sectional data refers to spatial units (municipalities, provinces, regions or countries) the spatial dependence between cross-sectional units at each point in time is also important. Spatial dependence implies that, due to spillover effects (e.g., commuter labour and trade flows), neighbouring regions may have similar economic performance. Hence, we expect to improve traditional panel data models by paying attention to the location of the spatial units. There has been growing interest in the estimation of static panel data models with spatial dependence: see Kelejian and Prucha (2002), Elhorst (2003), Yang et al. (2006), Baltagi et al. (2006), Kapoor et al. (2006), Kelejian et al. (2006) or Pesaran (2006). Prediction with these types of models is analysed in Baltagi and Li (2004, 2006) for predicting per-capita cigarette and liquor consumption in the United States, respectively, and in Longhi and Nijkamp (2007) for forecasting the regional labour market in West German regions. The extension of the traditional dynamic panel data model to include spatial effects has been worked on by Elhorst (2005) and Su and Yang (2007), who have derived the ML estimator of a dynamic panel data model extended to include cross-sectionally correlated error terms; Elhorst (2008), who derived the ML estimator for a spatially lagged dependent variable model (endogenous interaction effects); and Korniotis (2005) and Yu et al. (2007), who considered a dynamic panel data model extended to include both endogenous and lagged endogenous interaction effects.

The main objective of this paper is to evaluate the performance of the dynamic spatial panel data models in the forecasting of regional series. In our application, we use data on employment for 267 European regions (NUTS II administrative spatial unit in terms of Eurostat) from 1980 to 2008. The period 1980-2004 will allow us to estimate and check the models which, in a second step, will be used to forecast the series of employment by regions for the years 2005-2008. In this sense, our objective is similar to that of Kholodilin et al. (2008) when forecasting the GDP of German Länder. A novelty in our paper is that we evaluate the capacity of local estimation techniques for capturing any type of heterogeneity present in the best panel data specification.

The structure of the present paper is as follows. In Section 2, we provide a description of the panel data models we consider in our application. Section 3 is devoted to the presentation of the data and the main estimation results. In Section 4, we present the forecast performance of models. Finally, the paper finishes with a section of concluding remarks.

2. Panel data models

In this section, we describe a battery of models for panel data. We denote by R the number of spatial units (in our case, regions) we observe as cross-sectional data (i=1,2, …,R) and T denotes the total number of observations in the time dimension (t=1,2,…,T). Let’s start with the simple model, a *pooled* panel model:

$$\left.\begin{array}{l} y_t = x_t\beta + \eta_t \\ \eta_t \sim N\left[0, \sigma_\eta^2 I_R\right] \end{array}\right\} \tag{1}$$

$$\text{with} \quad y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ \vdots \\ y_{Rt} \end{bmatrix}; x_t = \begin{bmatrix} 1 & x_{21t} & \cdots & x_{k1t} \\ 1 & x_{22t} & \cdots & x_{k2t} \\ 1 & x_{23t} & \cdots & x_{k3t} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2Rt} & \cdots & x_{kRt} \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix}$$

which imposes the homogeneity restriction on both the intercept and slope coefficients across all the regions.

However, model (1) does not consider the probably presence of cross-sectional dependence among the observations at each point in time. To test this issue, Anselin et al. (2006) developed, in the context of panel data, the Lagrange Multiplier tests for a spatially lagged dependent variable and spatial error correlation, as well as their counterparts robustified versions. By considering the spatial interaction among observations we obtain the so-called *spatial panel data* models, which mainly adopt two forms: i) if we introduce the spatially lagged dependent variable as an explicative variable, we obtain the Spatial Lag Model (SLM) version defined as:

$$\left.\begin{array}{l} y_t = \rho W y_t + x_t\beta + \eta_t \\ \eta_t \sim N\left[0, \sigma_\eta^2 I_R\right] \end{array}\right\} \quad (2)$$

And, ii) if a spatially autorregresive process is incorporated into the error term, we obtain the Spatial Error Model (SEM), defined as:

$$\left.\begin{array}{l} y_t = x_t\beta + \varepsilon_t \\ \varepsilon_t = \rho W \varepsilon_t + \eta_t \\ \eta_t \sim N\left[0, \sigma_\eta^2 I_R\right] \end{array}\right\} \quad (3)$$

where W in equations (2) and (3) is the spatial weight matrix. As is well known, this matrix is pre-specified, nonnegative and of order RxR and describes the arrangement of the cross-sectional units in the sample (Anselin, 1988, 2007).

Models (2) and (3) consider the spatial interactions effects. However, they could also be improved when also considering the spatial specific effects, μ_i (i=1,2…,R), in order to account for the heterogeneity among spatial units. In fact, these terms represent the effect of omitted variables that are space-specific time-invariant variables that affect the dependent variable but are difficult to measure or hard to obtain. The following specifications are obtained:

SLM + Spatial specific effects	SEM + Spatial specific effects	
$\left.\begin{array}{l} y_t = \rho W y_t + x_t\beta + \mu + \eta_t \\ \eta_t \sim N\left[0, \sigma_\eta^2 I_R\right] \\ \mu = \left[\mu_1, \mu_2, ..., \mu_R\right]' \end{array}\right\}$	$\left.\begin{array}{l} y_t = x_t\beta + \mu + \varepsilon_t \\ \varepsilon_t = \rho W \varepsilon_t + \eta_t \\ \eta_t \sim N\left[0, \sigma_\eta^2 I_R\right] \\ \mu = \left[\mu_1, \mu_2, ..., \mu_R\right]' \end{array}\right\}$	(4)

The spatial specific effects may be treated as fixed effects or as random effects. In the fixed effects model, a dummy variable is introduced for each spatial unit, while in the random effect model, μ_i is treated as a random variable that is independently and identically distributed with zero mean and variance σ_μ^2. Furthermore, it is assumed that the random variable μ and ε_t are independent of each other. The random effects model can be tested against the fixed effects model using Hausman's specification test (Baltagi, 2005). If the hypothesis is rejected, the random effects models must be rejected in favour of the fixed effects model. However, discussion about random or fixed effects goes further than the only use of the Hausman's specification test. In the context of spatial data, the situation may be summarized according to two different positions.

On the one hand, models including a spatial structure need a very large sample (a large *R*, number of regions, in our case), because the convergence results are obtained with *R* tending to infinite. But, on the other hand, if the omitted effects are non-random, a problem of incidental parameters appears (the number of parameters grows at the same rate as the number of observations); in that case, a large *T* and small *R* are preferable. The last observation leads Anselin et al. (2006) to discard the use of fixed effects in mechanisms of spatial dependence: '*Since spatial models rely on asymptotics in the cross-sectional dimension (…), this would preclude the fixed effects model from being extended with a spatial lag or spatial error term*'. These authors prefer the random effect framework, where the inference is conditional and we only need a very large *R* (the improvements with T are of minor importance).

Elhorst (2003) does not share that view when he states that: '*The spatial units of observation should be representative of a larger population, and the number of units should potentially be able to go to infinity in a regular fashion. Moreover, the assumption of zero correlation between μ_r and the explanatory variables is particularly restrictive. Hence, the fixed effects model is compelling, even when R is large and T is small*'.

Further improvement of the models could be obtained by introducing the serially lagged dependent variable for capturing the inertia present in the temporal data or, in other words, for taking into account the serial dependence between observations on each cross-sectional unit over time. By doing this, we obtain the following dynamic specifications:

Dynamic SLM + Spatial specific effects	Dynamic SEM + Spatial specific effects
$\left.\begin{array}{l} y_t = \tau y_{t-1} + \rho W y_t + x_t\beta + \mu + \eta_t \\ \eta_t \sim N\left[0, \sigma_\eta^2 I_R\right] \\ \mu = \left[\mu_1, \mu_2, ..., \mu_R\right]' \end{array}\right\}$	$\left.\begin{array}{l} y_t = \tau y_{t-1} + x_t\beta + \mu + \varepsilon_t \\ \varepsilon_t = \rho W \varepsilon_t + \eta_t \\ \eta_t \sim N\left[0, \sigma_\eta^2 I_R\right] \\ \mu = \left[\mu_1, \mu_2, ..., \mu_R\right]' \end{array}\right\}$

(5)

Estimation of static spatial panel data models are explained by Elhorst (2003) and dynamic spatial panel data models can be estimated as explained by Elhorst (2005, 2008).

3. Data and estimation

In this application, we use data on employment for each of the 217 European regions (NUTS II administrative spatial unit in terms of Eurostat). Employment will be explained by investment and remuneration. The three variables are gathered for the period 1980 to 2008 from the Cambridge Database. The spatial distribution of the employment in the total economy in four different years appears in Figure 1.

(Insert Figure 1)

In order to define the proper specification for the models described in Section 2, we start by analyzing the statistical properties of data for each region, in logarithms (to reduce heterogeneity). The evolution of the log of employment in level and in its first differences, for each region, is displayed in Figures 2 and 3, respectively.

(Insert Figures 2 and 3)

The non-stationarity of the variables is confirmed by the applications of the common tests for unit roots in panel data. From the results obtained, shown in Table 1, we can conclude that the series are integrated of order one. As a consequence, the dependent variable for all panel data models will be the first differences of the logarithm of employment; that is, we will explain the growth rate of employment. The similar approach applied to the explicative variables reaches us to consider as explicative variables the growth rates of both variables, investment and remuneration. Furthermore, a temporal lag of such explicative variables will be included in order to avoid simultaneity problems.

(Insert Table 1)

In order to estimate spatial models, we must specify one or several weight matrices to reflect the network of cross-sectional relationships in the system of regions. To this respect, we have decided to develop a mixed neighbourhood criterion, which consists of the following. In the first place, we have used a criterion of neighbourhood based on the distance between the centroids of the regions. Furthermore, to avoid situations of excessive imbalance, we have opted to qualify the distance criterion by also incorporating the r nearest neighbours. Thus, we define the binary matrix W^b as:

$$w_{ij}^{b}(k,r) = \begin{cases} \text{if} \quad \min\limits_{i \neq s}\{d_{is}\} > k \quad \Rightarrow \begin{cases} w_{ij}^{b}(k,r) = 1 & \text{if} \quad j \in N_r(i) \\ w_{ij}^{b}(k,r) = 0 & \text{if} \quad j \notin N_r(i) \end{cases} \\ \\ \text{if} \quad \min\limits_{i \neq s}\{d_{is}\} \leq k \quad \Rightarrow \begin{cases} w_{ij}^{b}(k,r) = 1 & \text{if} \quad d_{ij} \leq k \\ w_{ij}^{b}(k,r) = 0 & \text{if} \quad d_{ij} > k \end{cases} \end{cases} \tag{6}$$

where d_{ij} is the distance in kilometres between the centroids of regions i and j, k the interaction radii and $N_r(i)$ the set of the r regions closest to region i. As usual, $w_{ii}=0$ for all i. In the paper we offer the results obtained for k=600 and r =2 but we have checked that results are very consistent with other values.

Firstly, we confirm the present of instability in the spatial cross-sectional model for several years by using the Lagrange Multipliers (LM) tests developed in Angulo et al. (2008) and Mur et al. (2009). The statistics refers to testing instability in the coefficient of spatial autocorrelation, the coefficients of the regression and/or the dispersion parameter, with the base of the following ample models:

For SLM type:

$$\left.\begin{array}{l} y = \rho \mathbf{HW} y + \hat{X}m + \varepsilon \Rightarrow \mathbf{A}y = \hat{X}m + \varepsilon \\ \varepsilon \sim N(0,\mathbf{\Omega}); \quad \mathbf{\Omega} = \sigma^2 \mathbf{D} \end{array}\right\}$$

For SEM type:

$$\left.\begin{array}{l} y = \hat{X}m + u \\ u = \rho \mathbf{HW} u + \varepsilon \Rightarrow \mathbf{B}u = \varepsilon \\ \varepsilon \sim N(0,\mathbf{\Omega}); \quad \mathbf{\Omega} = \sigma^2 \mathbf{D} \end{array}\right\} \tag{7}$$

Where

$$\left.\begin{array}{l} H = \operatorname{diag}\{h(z_i \alpha); i = 1, 2, ..., R\} \\ h(0) = \kappa \end{array}\right\} \tag{7a}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_R \end{bmatrix}_{Rx1}; \hat{X} = \begin{bmatrix} x'_1 & 0 & \cdots & 0 \\ 0 & x'_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & x'_R \end{bmatrix}_{RxRk}; x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \cdots \\ x_{ik} \end{bmatrix}_{kx1}; m = \begin{bmatrix} m_1 \\ m_2 \\ \cdots \\ m_R \end{bmatrix}_{Rkx1}; m_i = \begin{bmatrix} m_{i1} \\ m_{i2} \\ \cdots \\ m_{ik} \end{bmatrix}_{kx1} \quad (7b)$$

$$m_{ij} = \beta_j p_j \left[g'_i \mu \right] \begin{cases} g'_i = \left[g_{i1}; g_{i2}; \cdots; g_{ip} \right] \\ \mu'_i = \left[\mu_1; \mu_2; \cdots; \mu_p \right] \end{cases}$$

$$\left. \begin{array}{l} \varepsilon \sim N(0,\mathbf{\Omega}); \quad \mathbf{\Omega} = \sigma^2 \mathbf{D} \\ \Omega_{ii} = \sigma^2 d\left[n'_i \delta \right] \\ \Omega_{ij} = 0; i \neq j \\ d(0) = \kappa \end{array} \right\} \quad (7c)$$

Underlying (7a) is the hypothesis that the variability of the variable z generates instability in the coefficients of spatial dependence. Underlying (7b) is a situation of structural break in the regression coefficients assuming that the vector of coefficients changes at each point in space. Finally and underlying (7c), heteroskedasticity is introduced into the model assuming that it is generated by the variable n. On the base of the ample models defined above it is possible to derive all the battery of statistics described in Table 2. Results for all such tests together with the traditional Moran, non-robust and robust spatial aucorrelation LM test (Lmerror and Lmlag) and Sarma tests are gathered in Table 3. As expected, there are clear symptoms of instability in the three considered dimension: regression parameters, dispersion parameter and spatial dependence parameter. These results show that we are facing a problem that it is likely to be better modeled by means of panel data models, since the incorporation of unobservable heterogeneity of the data will reduce instabilities.

(Insert Tables 2 and 3)

The proper specification of spatial panel data model is derived through the estimation of alternative models. Results are gathered in Table 4.

(Insert Table 4)

Firstly, we estimate the pool model. However, according to the LM test for testing the null of no spatial effects (Anselin et al., 2006), the model suffers from misspecification, being a Spatial Lag Model the spatial structure that underline the data. The pooled SLM model (second column of results) outperforms previous model, but either the FE-SLM or the RE-SLM are better specifications. From Hausman’s test the RE-SLM model cannot be rejected. However, we support Elhorst’ s point of view, since we believe that the unobservable effects (or the omitted variables they are representing) are probably correlated with our explanatory variables. Moreover, FE-SLM model present a higher goodness of fit (R^2). Nevertheless, we will also compare the forecast performance of both models in next section.

Finally, we introduce temporal dynamic into the FE-SLM model. Results show that this model outperforms the previous one. However, a centre-periphery pattern can still be observed in the residual terms of the model[1]. As a consequence, a more flexible model is estimated by introducing a dummy variable (that takes the value of one for the periphery regions and zero value, otherwise) interacting with the two explicative variables. As deduced from results displayed in Table 4, this last model represent, till now, the best specification for our data set.

The question that we try to answer now it is to what respect all heterogeneity initially present in the data has been gathered with our flexible dynamic FE-SLM panel data model with centre-periphery interacting dummies. In order to solve this question, we propose a simple exercise consisting of using the local estimation technique with the reference model Dynamic FE+SLM. If results from local estimation technique outperform previous ones, we can conclude that spatial panel data models cannot capture all heterogeneity inherent in spatial data and therefore, they can be benefit from the use of local estimation techniques.

Briefly stated, the local estimation technique consists of fitting individual regressions to selected points in the sample, with more weight assigned to observations that are closer to the point of interest (McMillen 1996). Repeating this exercise for every point in the sample, we can construct estimation surfaces in order to discuss the nonstationarity of each parameter in the model. The concept of '*closeness*' is flexible and must be adapted to the objectives of the study. Moreover, the distribution of the weights among the neighboring observations with respect to point r is determined by the kernel function (Cressie 1991). In the case of the GWR, this is a decreasing function of the distance between the points, and the bandwidth determines how rapidly the weights decline with distance. We decided to use a rectangular or uniform kernel with a fixed bandwidth of *m* for every point. This means that the *m* nearest neighbors will receive a weight of one, and the other points zero.

[1] Results are available from the authors.

In our case we have to resolve the local estimation of an Dynamic FE+SLM for which it is not advisable to use the OLS algorithm. Following the example of Brunsdon et al. (1998) and of Pace and Lesage (2004), we will obtain the local estimators from the ML estimation of the local model:

$$\left.\begin{aligned} & y_{t,r}^{(m)} = \tau_r^{(m)} y_{t-1,r}^{(m)} + \rho_r^{(m)} W_r^{(m)} y_{t,r}^{(m)} + x_{t,r}^{(m)} \beta_r^{(m)} + u_r^{(m)} + \eta_{t,r}^{(m)} \\ & \eta_{t,r}^{(m)} \sim N\left[0, \sigma_{\eta,rm}^2 I_m\right] \\ & \mu_r^{(m)} = \left[\mu_1, \mu_2, \ldots, \mu_m\right]' \end{aligned}\right\} \qquad (8)$$

The indexes r and m mean that the data correspond to the local system defined by m elements around point r. Therefore, $y_{t,r}^{(m)} = (y_{t,r}, y_{t,i_1}, y_{t,i_2}, \ldots, y_{t,i_{m-1}})$ where $i_k \in N(r)$, being N(r) the bundle of indexes of the m-1 neighbours nearest to the point r. The same criterion is used to define $x_{t,r}^{(m)}$. Matrix $\mathbf{W_r^{(m)}}$ refers to the weighting matrix obtained for this local system, defined with the same connectivity criteria that are used to obtain the global W matrix, specified following standard criteria. Finally $\rho_r^{(m)}$, $\beta_r^{(m)}$ and $\sigma_{r,m}^2$ are the local parameters of interest. This is what we call the *Zoom* estimation (different to the SALE algorithm of Pace and Lesage, 2004, in that, in each local system, we use the matrix $\mathbf{W_r^{(m)}}$ specific for the local network around point r). We refer to *m* as the *Zoom size* (equivalent to *window size* in nonparametric literature).

Results for the estimated parameters through the Local estimation techniques are shown in Figure 4. As can be observed, there exist important differences among regions not only in magnitudes but also, in the cases of parameters associated to remuneration and investment, even in sign.

(Insert Figure 4)

Finally, in the next section, we analyze the forecast performance of all the spatial panel data models.

4. Forecast performance of the different models

The purpose of this section is to obtain and evaluate the employment forecast for all the regions for the period 2005-2008 derived from the different spatial panel data models.

Godberger (1962) shows that the best linear unbiased predictor (BLUP) for the cross-sectional units in a linear regression model $Y = X\beta + \eta$ with disturbance covariance matrix Ω at a future period T+C is given by:

$$\widehat{Y_{T+C}} = X_{T+C}\hat{\beta} + \Psi'\Omega^{-1}e \tag{9}$$

where $\Psi = E(\eta_{T+C}\eta)$ is the covariance between the future disturbance η_{T+C} and the sample disturbance η, X represents the independent variables of the model, $\hat{\beta}$ is the estimator of β, and e denotes the residual vector of the model. Elhorst (2009) derived the prediction formulas for the fixed effects and random effects model with a spatially lagged dependent variable.

Formulas for the fixed effect models are straightforward as $\Psi = 0$ provided that error terms are not serially correlated over time. Hence, predictions for the FE+SLM model can be derived as:

$$\widehat{Y_{T+C}} = \left(I_R - \hat{\rho}W\right)^{-1}\left(X_{T+C}\hat{\beta} + \hat{\mu}\right) \tag{10}$$

and prediction for all variants of such model can be derived by defining the X matrix accordingly.

Unlike the fixed effects model, the correction term $\Psi'\Omega^{-1}e$ in the random effect model is not zero. In the random effects spatial lag model, RE-SLM, predictions can be calculated as follows:

$$\widehat{Y_{T+C}} = \left(I_R - \hat{\rho}W\right)^{-1}\left[X_{T+C}\hat{\beta} + \left(1-\hat{\theta}^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} e_{1t} \\ \ldots \\ e_{Nt} \end{pmatrix}\right)\right] \tag{11}$$

where $\left(1-\hat{\theta}^2\right) = \dfrac{T\sigma_\mu^2}{T\sigma_\mu^2 + \sigma_\eta^2}$

That is, for the RE-SLM model to calculate the correction term $\Psi'\Omega^{-1}e$, the residual of each spatial unit are first averaged over the sample period and then multiplied with $\left(1-\hat{\theta}^2\right)$, a factor that can take values between 0 and 1.

Forecast performance of each spatial paned data model is evaluated through the Mean Absolute Percent Error (MAPE) measure, which for each time period t (forecast horizon equal to t-2004) is defined as:

$$MAPE_t = \frac{1}{N}\sum_{r=1}^{N}\frac{\left|employment_{r,t} - \widehat{employment}_{r,2004}(t-2004)\right|}{employment_{r,t}} * 100 \quad (12)$$

$$t = 2005 - 2008$$

Table 5 shows figures for the different MAPE quantitative magnitudes while Figure 5 shows the spatial distributions of the temporal mean of the absolute percent prediction errors.

(Insert Table 5 and Figure 5)

Results from Table 5 shows that predictions from the FE+SLM model clearly outperform that obtained from the RE+SLM model. Hence, these results confirm our intuitive decision on the preference of fixed effects over random effects. As regards as results for the other more ample FE+SLM models considered, we have conclude the following. Both, the dynamic FE+SLM model and its extensions (when we incorporate the structural break associated to the centre-periphery location of regions) offer better results in terms of the MAPE. However, the local estimation dynamic FE+SLM referred to spatial units outperforms previous results. Consequently, we can conclude that zoom-estimation also can be of help in the context of panel data in order to capture the remaining heterogeneity of models. Furthermore, we can observe that improvement is especially important as the prediction horizon increases.

Finally, we analyse the possibility of further improvement in predictions with the application of local estimation referred not only to the spatial unit but also to the time dimension. That is, whether or not we can better forecast a region when using the most recent observations in time of its neighboring spatial units. The obtained forecast results are shown in the last column of the Table (Dynamic FE+SLM+spatial and time zoom). Since for this last case the Mean Absolute Prediction Errors are larger than previous ones, we can conclude that, for this particular case, the use of the zoom estimation associated to the cross-sectional dimension of data is the best option.

5. Concluding remarks

Econometric literature clearly accepts the good performance of panel data models for being able to capture the unobservable heterogeneity of data. The empirical application offered in this paper has shown clear evidence on the fact that forecast results of flexible panel data models can still be improved by making use of local estimation techniques.

Further research will be directed towards the evaluation of the effect on results of certain decision about the size of the spatial or time zoom, as well as the possible effect of the selected functional form of the base model.

References

Angulo et al. (2008) An inquiry into the causes of instability in cross-sectional models: An application to outcomes in 2008 Spanish General Elections. 14th Internacional Conference on Computing in Economics and Finance, Paris.

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Anselin L, Le Gallo J, Jayet H (2006) Spatial panel econometrics. In: Matyas L, Sevestre P (eds) The econometrics of panel data, fundamentals and recent developments in theory and practice, 3rd edn. Kluwer, Dordrecht, pp 901-969.

Anselin L (2007) Spatial Econometrics. In: Mills TC, Patterson K (eds) Handbook of econometrics, Vol. 1. Palgrave, Basingstoke, pp 901-969

Arellano M (2003) Panel data econometrics. Oxford University Press, Oxford

Baltagi BH (2005) Econometric Analysis of Panel Data (3rd Edition). Wiley, Chichester

Baltagi BH, Li D (2004) Prediction in the panel data model with spatial correlation. In: Anselin L, Florax RJGM, Rey S (Eds) Advanced in spatial econometrics: methodology, tools and application. Springer-Verlag, Heidelberg (Germany), pp 283-295

Baltagi BH, Li D (2006) Prediction in the panel data model with spatial correlation: the case of liquor. Spatial Economic Analysis 1(2): 175-185

Baltagi BH, Song SH, Jung BC, Koh W (2006) Testing for serial correlation, spatial autocorrelation and random effects using panel data. Journal of Econometrics 140: 5-51

Brunsdom C, Fotheringham S, Charlton M (1998) Spatial Nonstationarity and Autoregresive Models. Environment and Planning A 30: 957-973

Cressie N (1991) Statistics for Spatial Data. Wiley, New York

Elhorst JP (2003) Specification and Estimation of Spatial Panel Data Models. International Regional Sciences Review 26: 244-268

Elhorst JP (2005) Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. Geographical Analysis 37: 62-83

Elhorst JP (2008) Estimation of dynamic panels with endogenous interaction effects. II Wold Conference of the Spatial Econometrics Association. New York

Elhorst JP (2009) Spatial Panel Data Models. In Fischer M.M., Getis A. (eds.) *Handbook of Applied Spatial Analysis*. Springer-Verlag, Berlin, Forthcoming.

Goldberger AS (1962) Best linear unbiased prediction in the generalized linear regression model. Journal of the American Statistical Association 57:369-375.

Hsiao C (2003) Analysis of panel data (2nd Edition). Cambridge University Press, Cambridge

Kapoor M, Kelejian HH, Prucha IR (2006) Panel data models with spatially correlated error components. Journal of Econometrics 140: 97-130

Kelejian HH, Prucha IR (2002) 2SLS and OLS in a spatial autoregressive model with equal spatial weights. Regional Science and Urban Economics 32: 691-707

Kelejian HH, Prucha IR, Yuzefovich Y (2006) Estimation problems in models with spatial weighting matrices which have blocks of equal elements. Journal of Regional Science 46: 507-515

Kholodilin KA, Siliverstovs B, Kooths S (2008). A dynamic panel data approach to the forecasting of the GDP of German Länder. Spatial Economic Analysis 3(2): 195-207

Korniotis GM (2005) A dynamic panel estimator with both fixed and spatial effects. http://www.uni-kiel.de/ifw/konfer/spatial/korniotis.pdf

Longhi S, Nijkamp P (2007) Forecasting regional labor market developments under spatial heterogeneity and spatial correlation. International Regional Science Review 30: 100-119

López F, Mur J, Angulo A (2009) Local estimation of spatial autocorrelation processes. In: Paez A, Le Gallo J, Buliung R, Dall'Erba S (Eds.) Progress in Spatial Analysis: Theory and Computation, and Thematic Applications. Springer, Berlin (forthcoming)

McMillen D (1996) One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach. Journal of Urban Economics 40: 100-124

Mur J, Lopez F, Angulo A (2008) Symptoms of instability in models of spatial dependence. Geographical Analysis 40: 189-211.

Mur J, Lopez F, Angulo A (2009) Testing the hypothesis of stability in spatial econometric models. Papers in Regional Science. Forthcoming

Pace K, Lesage J (2004) Spatial Autoregressive Local Estimation. In: Getis A, Mur J, Zoller H (eds) Spatial Econometrics and Spatial Statistics. Palgrave, London, pp 31-51

Pesaran MH (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74: 967-1012

Su L, Yang Z (2007) QML estimation of dynamic panel data models with spatial errors. http://fp.paceprojects.f9.co.uk/Su.pdf

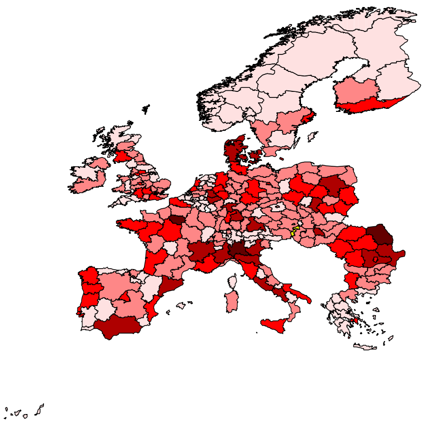
Wooldridge J (2002) Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge

Yang Z, Li C, Tse YK (2006) Functional form and spatial dependence in spatial panels. Economics Letters 91: 138-145

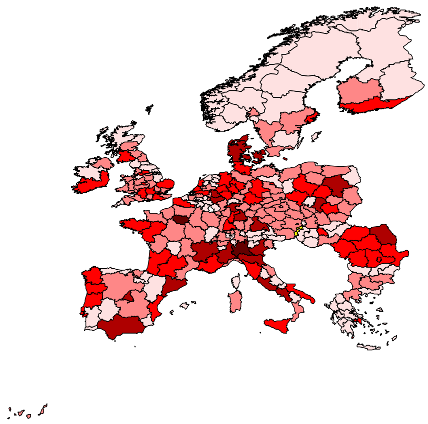
Yu J, de Jong R, Lee L (2007) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146(1):118-134

Figure 1. Spatial distribution of the employment

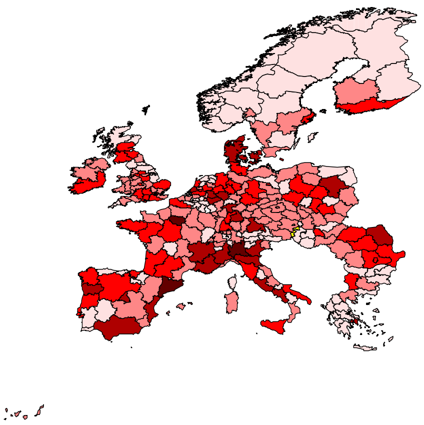
1980



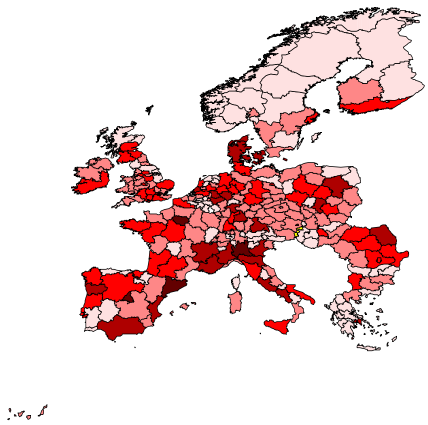
1995



2005



2008



In all cases:

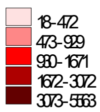


Figure 2. Evolution of the log(employment) in each of the 217 cross-sectional units (1980-2008)

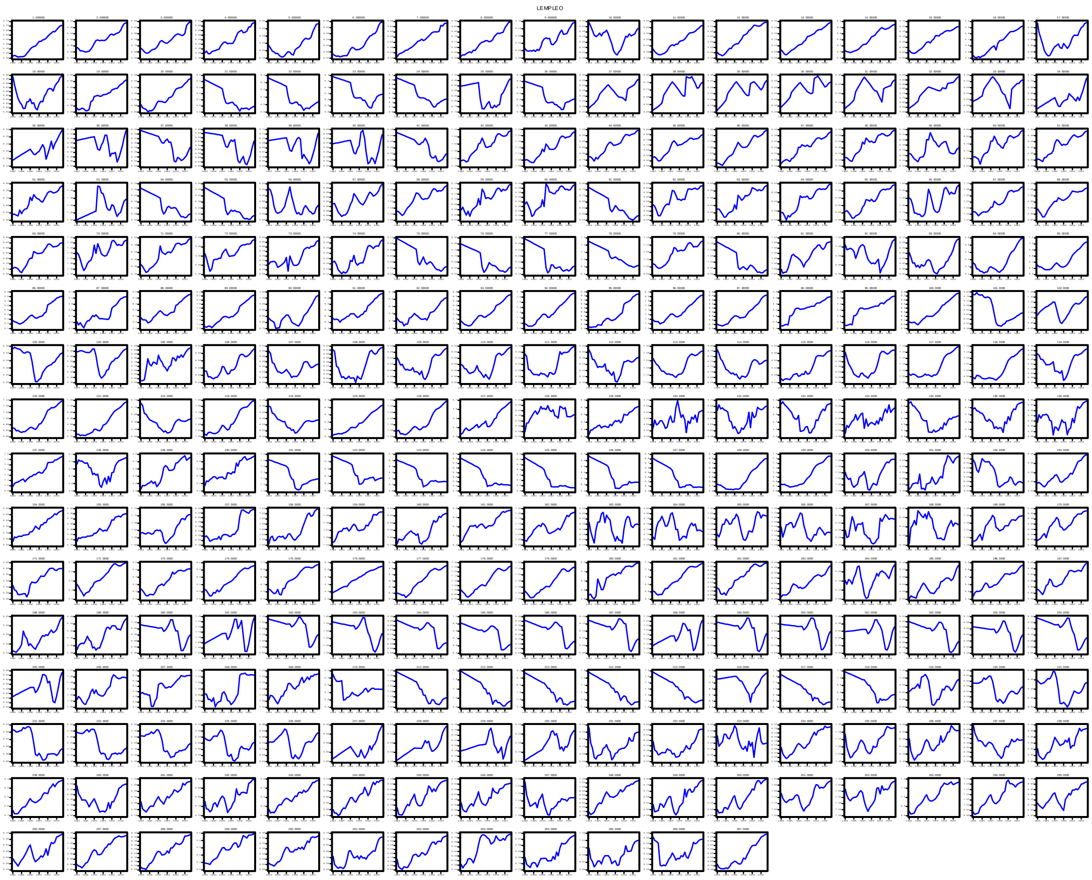


Figure 3. Evolution of the ∆log(employment) in each of the 217 cross-sectional units (1980-2008)

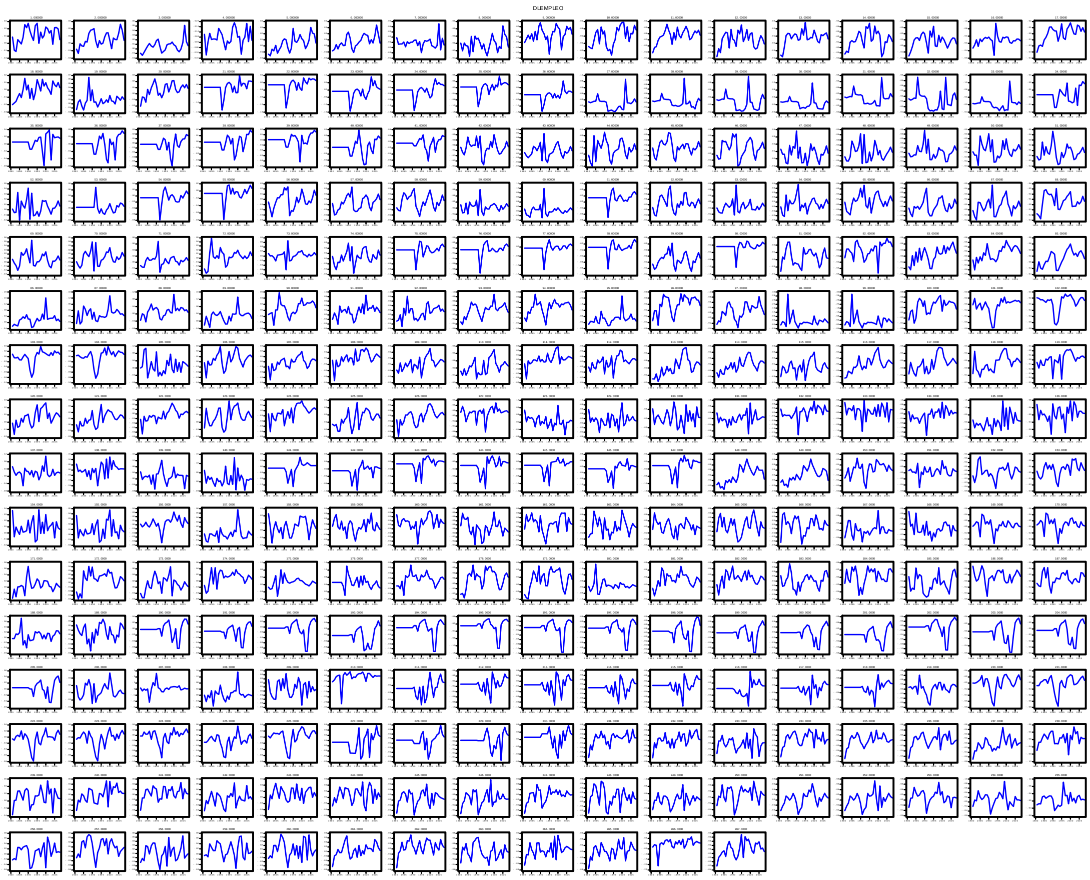


Figure 4. Spatial distribution of zoom parameters

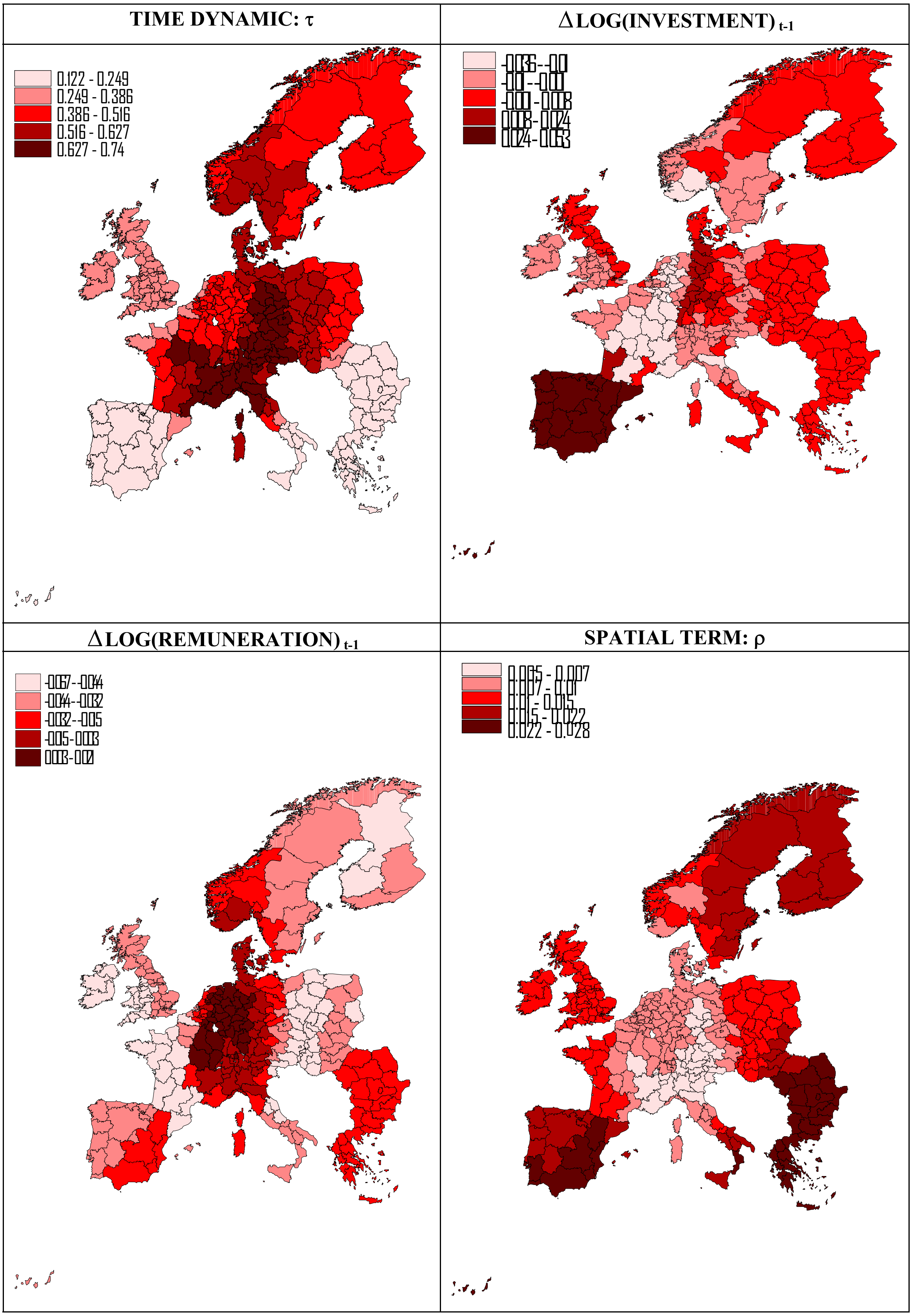


Figure 5. Spatial distribution of the temporal mean of the absolute percent predictions errors

Temporal Mean Absolute Percent Prediction Error

In all cases:

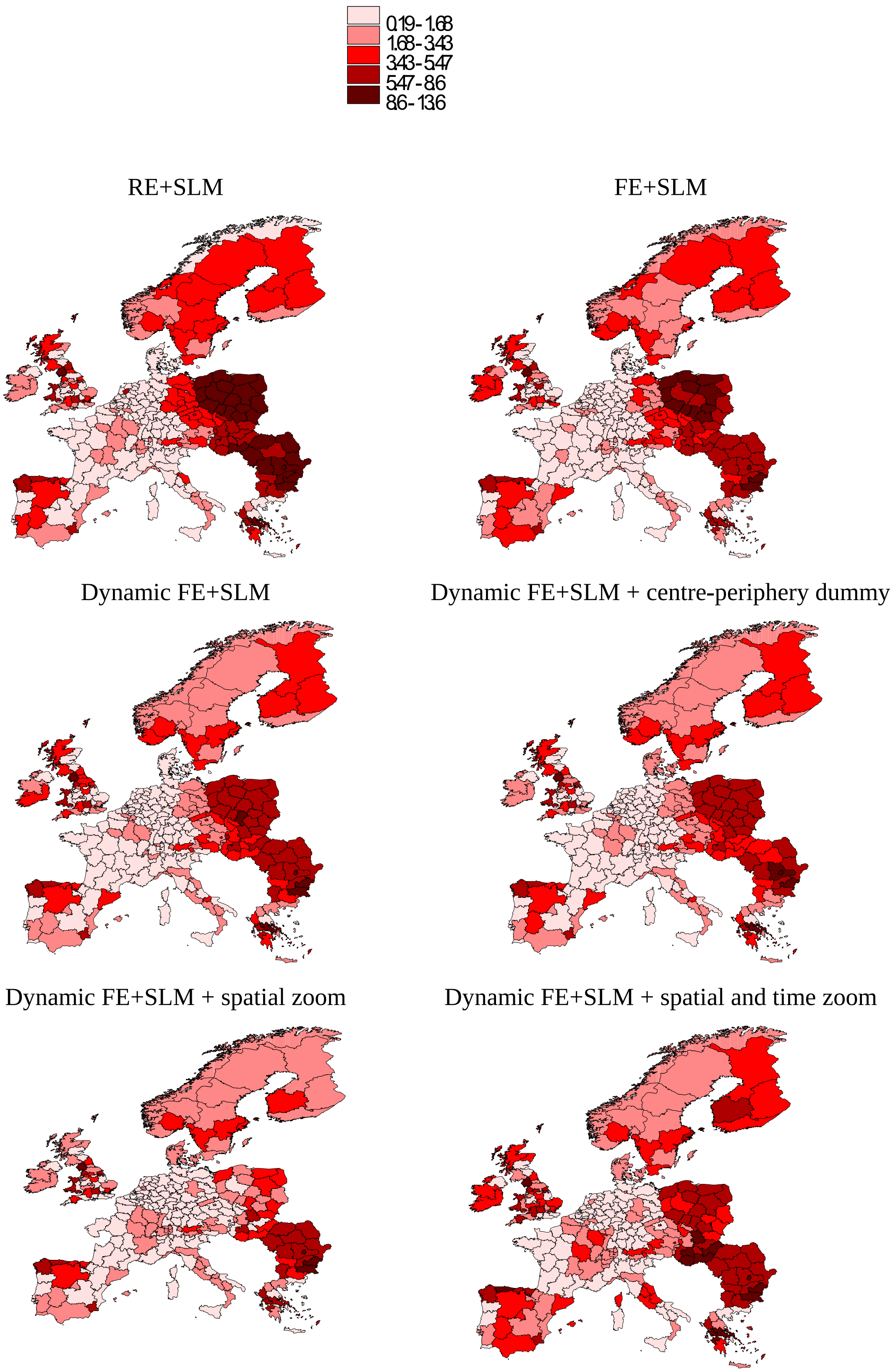


Table 1. Unit root test, under the assumption of specific mean and trend for each province

	Log (employment)		Log (investment)		Log (remuneration)	
	H_0: I(1) H_1: I(0)	**H_0: I(2) H_1: I(1)**	**H_0: I(1) H_1: I(0)**	**H_0: I(2) H_1: I(1)**	**H_0: I(1) H_1: I(0)**	**H_0: I(2) H_1: I(1)**
	Statistic (Prob.)	Statistic (Prob.)	Statistic (Prob.)	Statistic (Prob.)	Statistic (Prob.)	Statistic (Prob.)
H_0: Unit root (equal for all cross-sections)						
Levin, Lin & Chu (2002), t*	0.33 (0.63)	-23.09 (0.00)	-3.37 (0.00)	-27.67 (0.00)	-3.38 (0.00)	-33.9 (0.00)
Breitung	0.55 (0.71)	-17.39 (0.00)	-5.30 (0.00)	-28.77 (0.00)	-11.74 (0.00)	-24.36 (0.00)
H_0: Unit root (specific for each cross-sections)						
Im, Pesaran and Shin W-stat	7.83 (1.00)	-30.18 (0.00)	6.76 (1.00)	-37.42 (0.00)	2.76 (0.99)	-35.68 (0.00)
ADF - Fisher χ^2	360 (1.00)	1918 (0.00)	369 (1.00)	2403 (0.00)	492 (0.90)	2282 (0.00)
PP - Fisher χ^2	245 (1.00)	2917 (0.00)	316 (1.00)	4038 (0.00)	546 (0.35)	3528 (0.00)
Conclusion	**I(1)**		**I(1)**		**I(1)**	

Table 2. Robust Multipliers for testing the hypothesis of stability

Null Hypothesis	**Robust to**	**Statistic**	**Asymptotic Distribution**
Instability in σ^2, β and ρ H_0:$\delta=\mu=\alpha=0$		$\mathbf{LM}^{*SLM(SEM)}_{\delta,\mu,\alpha}$	$\chi^2(3)$
Instability in σ^2 H_0:$\delta=0$	Instability in ρ and β	$\mathbf{LM}^{*SLM(SEM)}_{\delta.(\mu,\alpha)}$	$\chi^2(1)$
Instability in β H_0:$\mu=0$	Instability in ρ and σ^2	$\mathbf{LM}^{*SLM(SEM)}_{\mu.(\delta,\alpha)}$	$\chi^2(1)$
Instability in ρ H_0:$\alpha=0$	Instability in σ^2 and β	$\mathbf{LM}^{*SLM(SEM)}_{\alpha.(\mu,\delta)}$	$\chi^2(1)$
Instability in σ^2 and β H_0:$\delta=\mu=0$	Instability in ρ	$\mathbf{LM}^{*SLM(SEM)}_{\delta\mu.(\alpha)}$	$\chi^2(2)$
Instability in σ^2 and ρ H_0:$\delta=\alpha=0$	Instability in β	$\mathbf{LM}^{*SLM(SEM)}_{\delta\alpha.(\mu)}$	$\chi^2(2)$
Instability in β and ρ H_0:$\mu=\alpha=0$	Instability in σ^2	$\mathbf{LM}^{*SLM(SEM)}_{\mu\alpha.(\delta)}$	$\chi^2(2)$

Table 3. Cross-sectional spatial dependence and instability tests

	1982	**1995**	**2000**	**2004**
Moran	6.84*	4.76*	10.67*	9.75*
Lmerr	34.38*	17.12*	92.16*	76.99*
Lmlag	25.33*	20.08*	98.44*	84.39*
Robust Lmerr	10.03*	0.68	0.06	0.07
Robust Lmlag	0.98	4.64*	6.35*	7.47*
Sarma	35.36*	20.76*	98.51*	84.46*
$\mathbf{LM}^{*SLM(SEM)}_{\delta,\mu,\alpha}$	32.88*	90.69*	645.69*	628.12*
$\mathbf{LM}^{*SLM(SEM)}_{\alpha.(\mu,\delta)}$	8.48*	15.89*	27.51*	32.38*
$\mathbf{LM}^{*SLM(SEM)}_{\delta.(\mu,\alpha)}$	17.97*	67.71*	542.34*	532.04*
$\mathbf{LM}^{*SLM(SEM)}_{\mu.(\delta,\alpha)}$	13.90*	22.76*	84.85*	82.03*
$\mathbf{LM}^{*SLM(SEM)}_{\delta\alpha.(\mu)}$	18.97*	67.93*	560.84*	546.09*
$\mathbf{LM}^{*SLM(SEM)}_{\mu\alpha.(\delta)}$	22.39*	38.66*	112.37*	114.42*
$\mathbf{LM}^{*SLM(SEM)}_{\delta\mu.(\alpha)}$	31.87*	90.47*	627.20*	614.07*

Table 4. Results obtained for the estimation of the different models

	POOL	SLM	RE+SLM	FE+SLM	Dynamic FE+SLM [a]	Dynamic FE+SLM With C-P interacting dummy [a]
CONSTANT	0.002* (4.89)	0.000 (0.33)	0.000 (0.15)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)
ΔLOG(INVESTMENT)$_{t-1}$						
All	0.037* (10.60)	0.016* (5.37)	0.015 (5.09)	0.015* (4.93)	0.018* (2.92)	
Centre						0.007 (0.95)
Periphery-Centre						0.043 (3.02)
ΔLOG(REMUNERATION)$_{t-1}$						
All	0.001 (0.17)	0.044 (1.18)	0.006 (1.78)	0.008* (2.32)	-0.004 (-0.54)	
Centre						-0.008 (-0.83)
Periphery-Centre						-0.001 (-0.06)
SPATIAL TERM: ρ		0.779* (51.79)	0.755 (46.61)	0.731* (42.22)	0.01* (11.53)	0.01* (11.53)
TIME DYNAMIC: τ					0.282* (11.03)	0.279* (10.89)
R^2	0.02	0.29	0.33	0.36	0.40	0.42
LM for no spatial error	3367.7*					
LM for no spatial lag	3500.4*					
Robust LM for no spatial error	21.89*					
Robust LM test for no spatial lag	154.64*					
LR for testing the null of no specific regional effects				633.3*		
Hausman Test			2.588			

[a] Dynamic panel data model must be estimated with a symmetric weight matrix, W. The range of ρ is in this case of (-0.091;0.017)

Table 5. Forecast performance of the different models, Mean of the Absolute Percent Error (MAPE), (%)

	RE+SLM	**FE+SLM**	**Dynamic FE+SLM**	**Dynamic FE+SLM With regional dummies**	**Dynamic FE+SLM + spatial zoom**	**Dynamic FE+SLM + spatial and time zoom**
MAPE(2005)	1.791	1.638	1.480	1.482	1.263	1.556
MAPE (2006)	2.954	2.696	2.512	2.437	1.974	2.726
MAPE (2007)	4.171	3.821	3.662	3.541	2.889	4.137
MAPE (2008)	4.937	4.323	4.155	4.006	3.356	5.126
MAPE	3.463	3.120	2.952	2.866	2.370	3.386

(a) MAPE(t) denote de mean of the absolute percent error for predicting the employment of period t in all the regions, $\text{MAPE(t)} = \frac{\sum_{i=1}^{N} \text{MAPE}_{\text{i}}\text{(t)}}{N}$. The term MAPE denotes the mean over the four predictions.