E C O N O M Í A

FONDO EDITORIAL

The Scan-LM to Test Instability in the Constant Coefficient of Spatial Autoregressive Models

Fernando A. López Hernández[a,*], Román Mínguez Salidos[b]

[a]Universidad Politécnica de Cartagena
  ✉ fernando.lopez@upct.es   *Corresponding author

[b]Universidad Castilla la Mancha
  ✉ roman.Minguez@uclm.es

Abstract

This paper presents a test based on the principle of Lagrange Multipliers to identify spatial instability in the constant coefficient of regression models including substantive spatial dependence. The test has been adapted to the Scan methodology. Its main advantage is that it identifies areas with differential behavior without the need to provide information about their location, shape, or size. The study shows the utility of the test, reconsidering the results obtained by Mur et al. (2008) about instability in the distribution of per capita income in European regions.

Acknowledgements

1.   Introduction

There is broad consensus that spatial dependence and heterogeneity are fundamental characteristics of spatial data (Griffith, 2003). The literature has largely focused on the former, spatial dependence, with numerous publications considering both testing and modeling issues. Spatial heterogeneity, however, has received less attention, even though there is considerable evidence that relationships tend to be unstable, making a degree of heterogeneity natural in cross-sectional models.

One of the possible sources of spatial heterogeneity is the presence of spatial structures in the model's coefficients. If this instability is not correctly included in the model, it can lead to spurious spatial autocorrelation (see, for instance, Lauridsen and Kosfeld, 2011). Assuming models have constant parameters is therefore somewhat unreal (Fotheringham et al., 1999). A large number of alternatives have been proposed to include instability in regression model parameters (such as Brunsdon et al., 1996; or, more recently, Lee et al., 2017), and several contributions include these instabilities in spatial regression models (for example, Basile et al., 2014 and Billé et al., 2017).

In general, two alternatives can be considered for including parametric instability in cross-sectional regression models: models in which parametric changes are continuous (such as Spline Regression Models) or discrete ruptures. This study is concerned with the latter. Only the recent contribution by Billé et al. (2017) is in the same line, proposing a two-step method that first identifies spatial regimes and then introduces spatial dependence.

The spatial heterogeneity problem in terms of the spatial variation of a model's parameters has been of interest in the spatial econometrics field. The literature contains several tests aimed at evaluating the stability of coefficients, largely based on the principle of Lagrange Multipliers (such as Lauridsen and Kosfeld, 2011) but, unfortunately, once the instability of coefficients is detected, there are no tools to suggest how to correctly model or identify which spatial regimens should be considered.

The starting point of our discussion is the paper by Mur et al. (2008), who considered an instability test in the spatial autocorrelation parameter. The authors maintain that the intensity of the occurrence of overflow effects between neighboring units cannot be considered constant. To illustrate how the test behaves, they contemplate a model that explains the per capita Gross Domestic Product (GDP) in European regions, identifying a spatial cluster of 611 regions (NUTS 3) in central Europe in which this intensity is greater than in the peripheral regions. Mentioned cluster is shown in Figure 1 (Figure 5 in Mur et al., 2008).

This study focuses on how the set of compact regions in central Europe was identified. The authors present evidence of the different behavior of spatial dependence structures, obtaining different Moran's I values for each country in the European Union (Mur et al., 2008, p. 195). In addition, they use an exploratory tool that they call Zoom Estimation to replicate the estimation of a model with local spatial effects for each region, considering only the closest regions (in terms of geographically weighted regressions, this would be the bandwidth). The different values of the spatial dependence parameter, ρ (called $\rho^{(i)}$), obtained using Zoom Estimation are charted on a
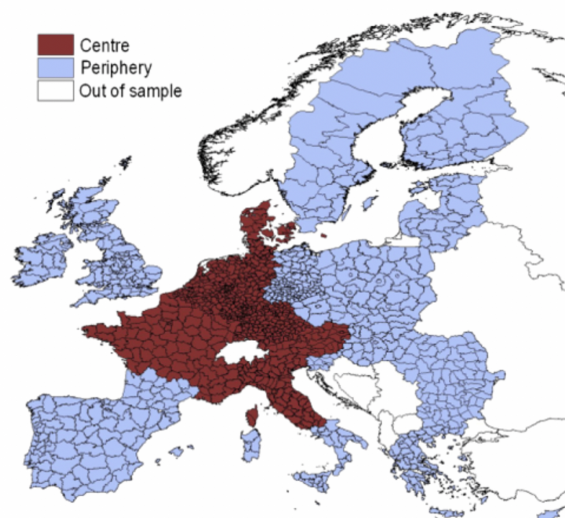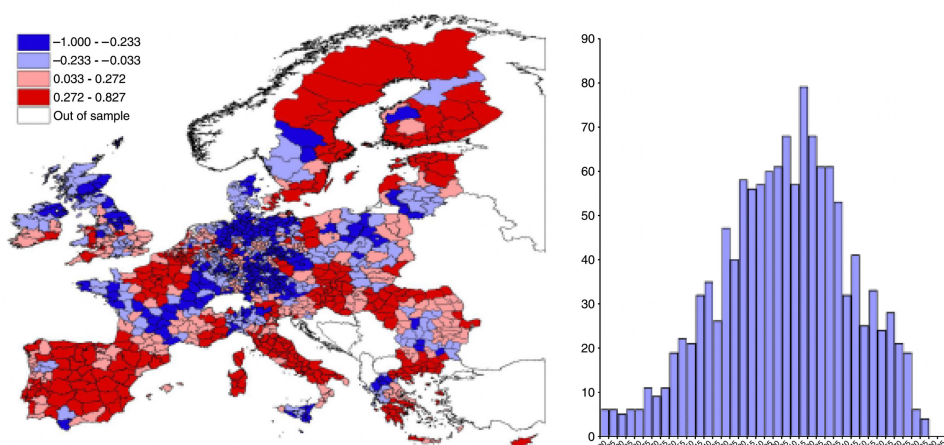
Source: Figure 5 in Mur et al. (2008), p. 199.

Figure 1. Center and periphery of European regions in Mur et al. (2008).

map. Figure 2 shows the map published in Mur et al. (2008), with all the $\rho^{(i)}$ values (Figure 3 in Mur et al., 2008).

This map constitutes the basis of decisions regarding the elements that configure the spatial cluster of differently behaving regions. Figure 1 shows the set of Central European regions with greater spatial dependence levels than the rest.

This is one more example of the methods based on **map observation** found in the literature to identify spatial clusters or areas with differential behavior. In other studies, it is also common to find the local Getis and Ord (1992) G_i^* statistic used to identify such clusters. Fischer and Stirböck (2006), for instance, or Le Gallo and Dall'Erba (2006), identify beta convergence clubs in European regions using this local indicator. Ramajo et al. (2008) also estimate a model with autocorrelation and spatial heterogeneity by selecting a set of regions not only guided by economic criteria but also considering the recommendation by Abreu et al. (2005), who suggest the use of Exploratory Spatial Data Analysis (ESDA) to select spatial regimes: "*In some sense, we follow the suggestion of Abreu et al. (2005) of basing the choice of regimes using ESDA information and a priori theoretical/political/geographical/economic considerations*"

This study contemplates a statistical methodology to identify clusters of regions/observations of differential behavior, free from the subjective criteria used in previous studies. The basis is the Scan statistic (see, for instance, Kulldorff and Nagarwalla, 1995), which has been used in spatial econometrics by various authors (López et al., 2015; Chasco et al., 2018; Le Gallo et al., 2020; Cucala, 2016). Our proposal focuses on the identification of spatial clusters in the mean of a spatial process specifically when instability lies in the independent term of the regression equation. This methodology can easily be used to identify instability in the coefficients of explanatory variables.

Figure 2. Center and periphery of European regions in Mur et al. (2008).

2.   The Scan-LM$_\alpha$ Test

We start with the specification of a SAR (Simultaneous Autoregressive Regression) autoregressive spatial model:

$$y = X\beta + \rho W y + u,$$

where β is a vector $(kx1)$ of the coefficients of the explanatory variables X (nxk); W is the classic connection matrix (nxn); ρ is the spatial autocorrelation parameter and we assume that the residuals are homoscedastic and normal: $u \sim N(0, \sigma^2 I_n)$.

Based on the alternative hypothesis, we contemplate instability in the equation's constant[1],

$$y = X\beta + Z\theta + \rho W y + u,$$

where Z is a variable with a value of 1 for some observations and 0 for the rest. The initial test considered is:

$$H_0 : \theta = 0$$
$$H_A : \theta \neq 0.$$

The general expression of the test based on the Lagrange Multipliers LM is (the test is shown in detail in Appendix A):

$$LM_\alpha = \left[g\left(\theta\right)\big|_{H_0} \right]' \left[\mathrm{MI}\left(\theta\right)\big|_{H_0} \right]^{-1} \left[g\left(\theta\right)\big|_{H_0} \right] \sim \chi_1^2,$$

where g is the gradient of the model's likelihood function based on the null hypothesis and MI is the information matrix.

---

[1]In this case, we assume that the first column of X is the unit vector. It can easily be applied to instabilities in the explanatory variables.

To be computationally efficient, it is essential to find a closed expression that enables us to distinguish between terms that depend on Z and terms that do not. The final expression after some calculations (see Appendix A) is:

$$LM_\alpha = \frac{1}{\sigma^2} Z'u \left[Z'Z - b'A^{-1}b \right] u'Z.$$

See Appendix A for a full definition of terms.

In this situation, in order to perform the test, the investigator needs to first formulate a hypothesis about the elements forming the cluster, defining Z a priori. The investigator's choice of Z is critical, as a test with several alternative definitions of Z could be affected by the multiple comparison problem.

The alternative proposed here is to adapt the test to the Scan methodology, proposing these hypotheses:

$$H_0 : \theta = 0 \left(\forall \Theta_Z \in \Theta \right)$$
$$H_A : \exists \Theta_Z \in \Theta \text{ where } \theta \neq 0 \left(\text{or } \theta > 0; \text{or } \theta < 0 \right),$$

where Θ is the set of all possible connected regions that can be considered in the study area and Θ_Z is the set of connected regions selected in the dummy variable Z. For practical purposes, this set Θ is usually limited to circular and/or elliptical regions, although flexibly shaped spatial clusters can also be considered (Tango and Takahashi, 2005).

The statistic proposed to test this hypothesis is:

$$\text{Scan-LM}_\alpha = \max_{\Theta_Z \in \Theta} \{LM_\alpha\}.$$

3. Significance of the Statistic

The theoretical distribution of the Scan statistic based on the null hypothesis is unknown. Resampling techniques are therefore required to evaluate its significance. This resampling method will provide a $p_R-value$ comparing the value of the statistic for the set of real data with the empirical distribution of R spatial reorders of the sample. The process is as described below:

1. Calculate the Scan-LM$_\alpha$ statistic for the original data set $\{y_i, X_i\}_{i \in S}$, where S is a set of spatial locations determined by their latitudinal and longitudinal coordinates.

2. Reorder the set of observations by permutation, randomly assigning observations to locations, and thus obtaining $\{y_i^r, X_i^r\}_{i \in S}$ where r is the index of the replica.

3. Perform the same permutation on the rows and columns of the W matrix, obtaining W^r, thus breaking the structure of Z as a cluster, while maintaining exactly the same spatial dependence structure.

4. Perform the Scan-LM$(r_i)_\alpha$ test on the permuted sample $\{y_i^r, X_i^r\}_{i \in S}$ with matrix W^r.

5. Repeat steps 2 and 4 $(R-1)$ times to obtain the statistic a total of R times $\{\text{Scan-LM}(r_i)_\alpha\}_{i=1}^{R}$.

6. Calculate the p-pseudo value of the statistic as:

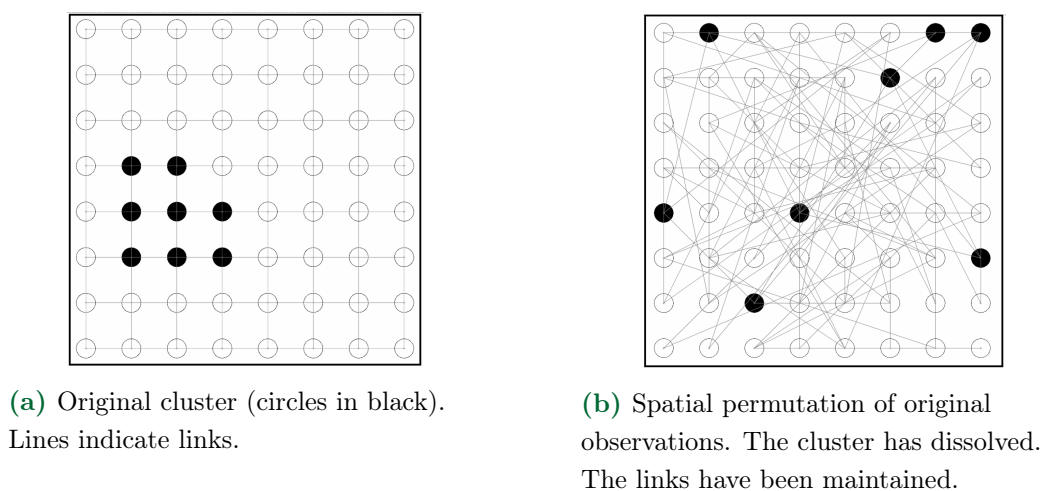$$p_R-value = \frac{1}{R} \sum_{i=1}^{R-1} I(\text{Scan-LM}_\alpha > \text{Scan-LM}(r_i)_\alpha),$$

where $I(\cdot)$ is an indicating function that assigns a value of 1 if inequality is true and 0 otherwise.

Reject the null hypothesis if $p_B-value < \alpha$ for a nominal value α. See Le Gallo et al. (2020) for more details about this procedure.

3.1   A Graphic Example of the Spatial Bootstrap

Figures 3a, 3b graphically show the resampling method described in Section 3. In Figure 3a, the Scan test identifies a spatial cluster of 8 observations with differential behavior (in black). The lines represent the system of connections defined by W. Through the scanning procedure, a total of $\#(\theta)$ dummy variables Z (windows) are considered, covering all the possible clusters of observations. The LM_α statistic should present the maximum value when the dummy variables identify those 8 observations. The significance of the statistic is evaluated by means of a random permutation of the observations (sampling without replacement) and the vector of explanatory variables X and matrix W. Figure 3b shows the spatially permuted observations y_p, and the lines show the new configuration of matrix W once permuted, W_p. Note that the rows and columns of matrix W are permuted in the same order, thus maintaining the initially established neighborhood criterion. The scanning processes are again initiated for this permutation, using all the windows predefined in θ but, in this case, a cluster is not identified and the values of the LM_α statistic are (in general) lower than those obtained with the original data. Note that the estimation of the SAR model has not varied as only the order of the observations has changed with the same permutation used for the vectors y, Xs, and matrix W.

Numerically speaking, the problem is the same as calculating the maximum of the LM_α tests

(a) Original cluster (circles in black). Lines indicate links.

(b) Spatial permutation of original observations. The cluster has dissolved. The links have been maintained.

Figure 3. Permutation resampling in a spatial environment.

for the connected windows with the distribution of the maximum LM_α statistic for B sets of unconnected windows.

3.2 Secondary Clusters

If the test rejects the null hypothesis and identifies a significant cluster, a natural question would be to ask if there is another cluster, not overlapping the Most Likely Cluster (MLC). These clusters are the so-called secondary clusters. Zhang et al. (2010) suggest an iterative method based on eliminating the observations included in the MLC from the sample and re-obtaining the value of the statistic with this subsample. Zhang et al. (2010) confirm that this procedure offers more power to identify secondary clusters. This will, therefore, be the method used in this paper.

4. The Reformulated Application of Mur et al. (2008)

In this section we will show the use of the Scan-LM_α statistic together with other tests based on the Scan methodology (López et al., 2015; Chasco et al., 2018; Le Gallo et al., 2020). The objective is only to show how this methodology can be used to ensure the correct specification of the econometric model with spatial effects.

The proposed econometric model is as contemplated in Mur et al. (2008), estimating the per capita GDP of the European regions in 1998 using two explanatory factors: population density ($dens$) and the percentage weight of the primary sector in the GDP (wag). It considers a total of 1,274 regions corresponding to NUTS 3 in the 27 European Union countries (now 28). The contact matrix W that encodes the relations among regions is based on a mixed criterion (distance between centroids and the closest neighbor). Additional information about the design of the model can be found in Mur et al. (2008).

The first model estimated in this section is a simple estimation using OLS.

$$\textbf{Model 1: } y = \beta_0 + \beta_1\, dens + \beta_2\, wag + u; \; u \sim N(0, \sigma^2 I).$$

The results of the estimation are shown in Table 1 (first column). The residuals of the OLS model show clear symptoms of presenting spatial structure. All the spatial autocorrelation tests (MI and ML) broadly reject the null hypothesis of independence. It is also rejected by the Scan_σ test (Chasco et al., 2018) and both effects (SGHW and spatial autocorrelation) are probably present in the model. The Scan_σ test provides additional information (Figure 4) identifying two clusters, one with high variance (214 regions, in red) and one with low variance (626 regions, in green). The specification clearly requires improvement.

We decided to solve the presence of spatial autocorrelation by including a spatial challenge to the endogenous autocorrelation. The SAR model with the matrix W described in Mur et al. (2008) has the following formulation:

$$\textbf{Model 2: } y = \beta_0 + \rho W y + \beta_1\, dens + \beta_2\, wag + u; \; u \sim N(0, \sigma^2 I).$$
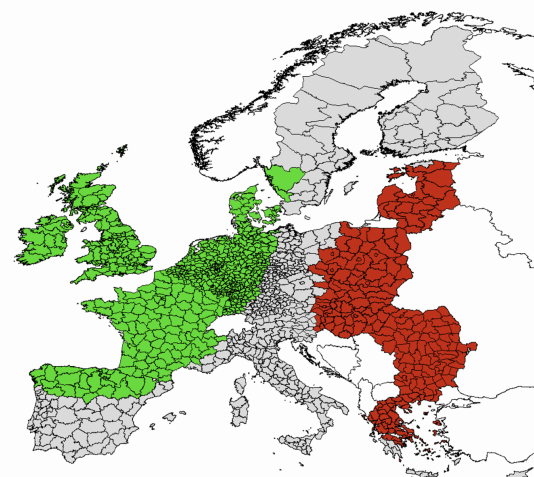
Source: own construction using QGIS.

Figure 4. Clusters identified in the residuals of Model 1.

The second column of Table 1 shows the results of the estimation. The spatial autocorrelation parameter presents a high value $\rho = 0.645$ but the diagnostic tests still show specification problems. The marginal $\mathrm{LM}_{\lambda|\rho}$ test rejects the null hypothesis, showing that the spatial dependence that the residuals of Model 2 presented remain unsolved and that including the term Wy was insufficient. There are also symptoms of instability in ρ as $\mathrm{LM}_{\mathrm{Break}}^{\mathrm{SAR}}(\mathrm{cb}) = 34.91$ rejects the null hypothesis of homogeneity in the parameter of spatial dependence.

The null hypothesis is also rejected by the Scan-$\mathrm{LM}_{\mathrm{SAR}}^{\mathrm{SGWH}}$ test (Le Gallo et al., 2020), showing that there are symptoms of Spatial GroupWise Heteroskedasticity (SGWH) in the residuals of the SAR model. The additional information provided by this test about the patterns of SGWH is shown in Figure 5. There are two clusters, one with maximum variance (198 regions, in red) and one with low variance (229 regions, in green). It is clear that the inclusion of spatial effects has reduced the spatial structure problems in the residuals of the model.

This could be affected by several issues: (i) incorrect selection of Xs; (ii) incorrect selection of matrix W; (iii) SGWH problems; (iv) instability in the spatial autocorrelation.

The Scan-LM_α test suggests issue (i), showing instability problems in the constant of the equation. The cause or causes of the problem cannot be identified at this point.

The Scan-LM_α test suggests a total of 6 ruptures in the intercept of the model (parameter α), (see Figure 6). With this information, we decided to include the respective dummy variables suggested by the Scan-LM_α test to correct the instability problems in the mean of the process. We define the matrix $\mathbf{Z}$ of dimension $(nx6)$, where each column (named Z_i) is a dummy variable to identify each cluster. The elements of the first column of the matrix $\mathbf{Z}$ (named Z_1) is a dummy variable to identify the regions included in the Most Likely cluster; Z_2 is the dummy variable to identify the regions selected in the first secondary cluster (see subsection 3.2); and Z_3 is the dummy variable to identify the regions in the second secondary cluster. The same procedure is used to define Z_4, Z_5, and Z_6. With the information obtained using the Scan-LM_α test, the
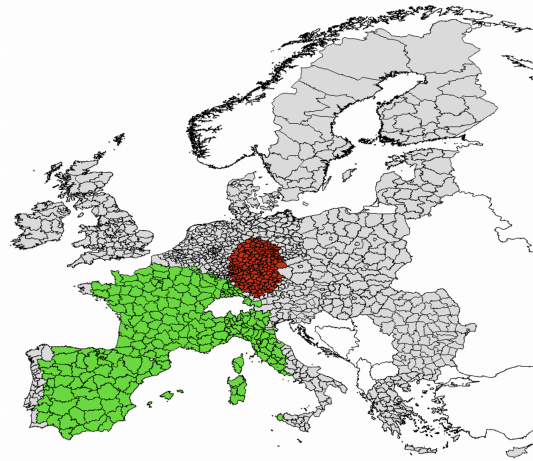
Source: own construction using QGIS.

Figure 5. Clusters identified in the residuals of Model 2.
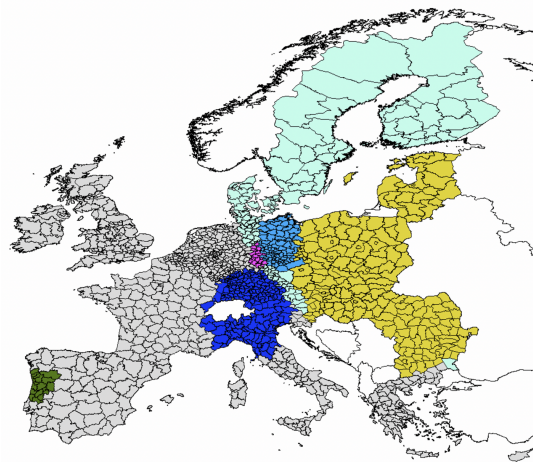
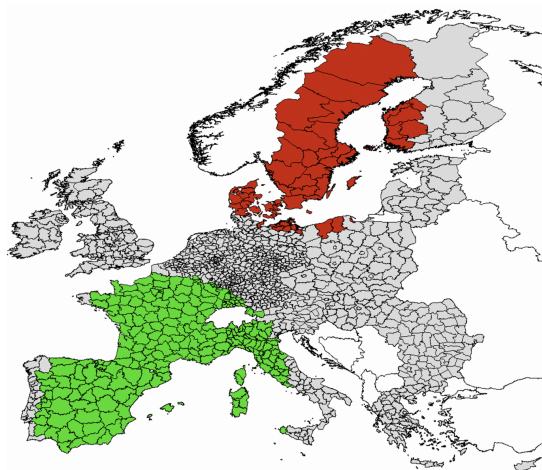*Source*: own construction using QGIS.

Figure 6. Ruptures suggested by Scan-LM$_\alpha$.

following specification is estimated:

$$\textbf{Model 3:}\ y = \beta_0 + \sum_{k=1}^{6} \theta_k Z_k + \rho W y + \beta_1 dens + \beta_2 wag + u;\ u \sim N(0, \sigma^2 I),$$

where Z_k is the dummy variable with a value of 1 if the region is in cluster k and 0 otherwise.

The third column of Table 1 shows the results of the estimation of Model 3. The most significant result is the drastic drop in the spatial dependence coefficient, which now has a value of $\rho = 0.378$, showing that part of the spatial structure identified in Model 2 was probably due to instabilities of the mean. This result confirms the statement published by Billé et al. (2017, p. 3), "*the spatial heterogeneity might generate part of the spatial autocorrelation effect, or in other words, that the autoregressive coefficient is sometimes overestimated*". The $LM_{\lambda|\rho}$ test continues to show that part of the spatial dependence has not yet been included in the model, although

PUCP

Source: own construction using QGIS.

Figure 7. Clusters identified in the residuals of Model 3.

with a lower value of the statistic (28.06 in Model 3 *vs.* 119.52 in Model 2). The $\text{LM}^{\text{SAR}}_{\text{Break}}$(cb) = 11.09 test also shows instability in the spatial autocorrelation parameters, although with a lower value than that obtained for the residuals of Model 2 (11.09 *vs.* 34.91). The Scan-$\text{LM}^{\text{SGWH}}_{\text{SAR}}$ test also rejects the null hypothesis, showing the possible presence of SGWH.

We again make use of the information provided by the Scan-$\text{LM}^{\text{SGWH}}_{\text{SAR}}$ test, now identifying (see Figure 7) a cluster with low incidence (235 regions, in green) and a small cluster (57 regions, in red).

Finally, following Mur et al. (2008), we decided to model the instability in the spatial dependence parameters. Model 4 is expressed as:

$$\text{Model 4:} \quad y = \beta_0 + \sum_{k=1}^{6} \theta_k Z_k + \rho W y + \gamma W^* y + \beta_1 \, dens + \beta_2 \, wag + u; \, u \sim N(0, \sigma^2 I).$$

We will use matrix W^* to identify the set of regions with peculiarities measured by the parameter γ. The W^* matrix must exactly reproduce a certain part of matrix W (the rows and columns associated with the case that we want to test, while the other elements are zero). In this case, the rupture area is the same as presented in Mur et al. (2008). The last column of Table 1 shows the results of the estimation of the model. The two spatial dependence parameters are significant, and Model 4 improves on Model 2 in terms of likelihood.

Finally, the Scan$_\sigma$ test is used on the residuals to evaluate the reduction in instability. Figure 8 shows the test's output. Only a low variance cluster (142 observations, in green) is identified in this case. It is clear that the instability in ρ was the cause of part of the problems present in the previous models.

An optimal solution appears to be out of reach, as it is always possible to identify a degree of instability in the residuals.
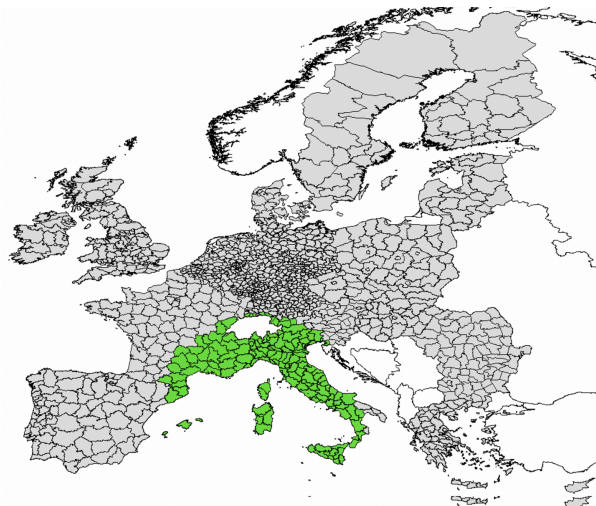
Source: own construction using QGIS.

Figure 8. Clusters identified in the residuals of Model 4.

Table 1

Results of the estimation.

	Model 1	Model 2	Model 3	Model 4
β_0	9.820***	3.470***	6.083***	6.258***
β_1 (*dens*)	0.075***	0.089***	0.092***	0.093***
β_2 (*wag*)	−0.050***	−0.021***	−0.022***	−0.022***
θ_1	–	–	−0.341***	−0.338***
θ_2	–	–	−0.186***	−0.165***
θ_3	–	–	0.134***	0.111***
θ_4	–	–	−0.247***	−0.227***
θ_5	–	–	−0.207***	−0.208***
θ_6	–	–	0.109***	0.102***
ρ	–	0.645***	0.378***	0.356***
γ	–	–	–	0.006***
Log Likelihood		−92.92	26.60	32.63
		Diagnostic tests		
$LM_{\gamma\mid\rho}$		119.52***	28.06***	–
$Scan_\sigma$	234.67***	–	–	32.49***
Scan-LM$_{\text{SAR}}^{\text{SGWH}}$	–	123.44***	47.12***	–
Scan-LM$_\alpha$	–	64.58***	–	–
Scan-LM$_{\text{Break}}^{\text{Lag}}$(cb[1])	–	34.91***	11.09***	–
I Moran	29.11***	–	–	–
LM-err	834.98***			
LM-EL	174.57***			
LM-lag	739.50***			
LM-LE	79.09***			
LM-SARMA	914.07***			

Notes: Scan$_\sigma$ is the Scan test for the residual (Chasco et al., 2018);
Scan-LM$_{\text{SAR}}^{\text{SGWH}}$ is the Scan test of SGWH (Le Gallo et al., 2020);
Scan-LM$_{\text{Break}}^{\text{Lag}}$ is the LM test for instability of spatial dependence in
SAR (Mur et al., 2008).

[1]cb is the area identified in Mur et al. (2008) formed by 611 regions.

5.   Conclusions

It is our impression that both types of structure (spatial autocorrelation and spatial heteroskedasticity) must be tested for correct specification. This information, moreover, is vital to improve the specification of the model, so a combination of spatial autocorrelation tests and SGWH must be used in any exercise modeling cross-sectional data (Chasco et al., 2018). The way in which these tests should be used is to be approached in future research.

The Scan methodology is widely used in a number of fields, and it has recently been incorporated into the area of spatial econometrics (López et al., 2015; Chasco et al., 2018, 2020). Compared to other contrast methodologies, the main advantage of the Scan methodology is its subproduct. Not only does statistically contrast the hypothesis about the instability parameter but it also identifies the spatial clusters of differential behavior. This second aspect is completely new since it substitutes decisions made by the researcher. In any case, the information provided by the contrast identifying the cluster(s) can always complement researchers' intuition when identifying structural changes.

This work has been developed only to identify structural changes in the constant of SAR spatial regression models. Amplifications to identify structural changes in the other types of spatial regression models are obvious and will occupy a part of our research agenda in the coming years.

Appendix A - Model

For the derivation of the test based on the Lagrange Multipliers, the first derivatives are necessary,

$$\frac{\partial 1}{\partial \beta} = \frac{1}{\sigma^2}(By - X\beta - Z\theta)'X = \frac{1}{\sigma^2}u'X$$

$$\frac{\partial 1}{\partial \theta} = \frac{1}{\sigma^2}(By - X\beta - Z\theta)'Z = \frac{1}{\sigma^2}u'Z$$

$$\frac{\partial 1}{\partial \rho} = -trB^{-1}W + \frac{1}{\sigma^2}(By - X\beta - Z\theta)'Wy = -trB^{-1}W + \frac{1}{\sigma^2}u'Wy$$

$$\frac{\partial 1}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^4} - \frac{1}{2\sigma^6}u'u$$

where $B = (I - \rho W)$, and with the gradient under the null hypothesis being

$$g(\beta, \theta, \rho, \sigma^2)\big|_{H_0} = \begin{bmatrix} \dfrac{\partial 1}{\partial \beta} \\[2mm] \dfrac{\partial 1}{\partial \theta} \\[2mm] \dfrac{\partial 1}{\partial \rho} \\[2mm] \dfrac{\partial 1}{\partial \sigma^2} \end{bmatrix}\bigg|_{H_0} = \frac{1}{\sigma^2}\begin{bmatrix} 0 \\ u'Z \\ 0 \\ 0 \end{bmatrix}.$$

The information matrix is:

$$
\mathrm{MI}\big|_{H_0} = \begin{bmatrix}
\dfrac{\partial 1}{\partial\beta\partial\beta} & \dfrac{\partial 1}{\partial\beta\partial\theta} & \dfrac{\partial 1}{\partial\beta\partial\rho} & \dfrac{\partial 1}{\partial\beta\partial\sigma^2} \\[2ex]
-- & \dfrac{\partial 1}{\partial\theta\partial\theta} & \dfrac{\partial 1}{\partial\theta\partial\rho} & \dfrac{\partial 1}{\partial\theta\partial\sigma^2} \\[2ex]
-- & -- & \dfrac{\partial 1}{\partial\rho\partial\rho} & \dfrac{\partial 1}{\partial\rho\partial\sigma^2} \\[2ex]
-- & -- & -- & \dfrac{\partial 1}{\partial\sigma^2\partial\sigma^2}
\end{bmatrix}\Bigg|_{H_0}
$$

In this case, we can express it as:

$$
\mathrm{MI}\big|_{H_0} = \frac{1}{\sigma^2}\begin{bmatrix}
X'X & X'Z & ((B^{-1}X\beta)'WX)' & 0 \\
-- & Z'Z & (B^{-1}X\beta)'WZ & 0 \\
-- & -- & \sigma^2 trB^{-1}WB^{-1}W+ & trWB^{-1} \\
 & & +[\beta'X'W'WX\beta + \sigma^2 trW'W] & \\
-- & -- & -- & \dfrac{n}{\sigma^2}
\end{bmatrix}.
$$

The general expression for the LM test is:

$$
\text{Scan-LM}_\alpha = \left[g|_{H_0}\right]'\left[\mathrm{MI}\big|_{H_0}\right]^{-1}\left[g|_{H_0}\right],
$$

that after some calculations, permuting the fourth and second rows of the gradient and the second and fourth rows and columns of the information matrix and taking into account the expression of the inverse of a block matrix, we get:

$$
\frac{1}{\sigma^2}\begin{bmatrix}0 & 0 & 0 & Z'u\end{bmatrix}\begin{bmatrix}
X'X & 0 & ((B^{-1}X\beta)'WX)' & X'Z \\
0 & \dfrac{n}{\sigma^2} & trWB^{-1} & 0 \\
(B^{-1}X\beta)'WX & trWB^{-1} & \sigma^2 trB^{-1}WB^{-1}W+ & ((B^{-1}X\beta)'WZ)' \\
 & & +[\beta'X'W'WX\beta + \sigma^2 trW'W] & \\
Z'X & 0 & (B^{-1}X\beta)'WZ & Z'Z
\end{bmatrix}\begin{bmatrix}0\\0\\0\\u'Z\end{bmatrix}
$$

and

$$
\text{Scan-LM}_\alpha = \frac{1}{\sigma^2}Z'u\left[Z'Z - b'A^{-1}b\right]u'Z,
$$

with

$$
b = \begin{bmatrix}X'Z\\0\\((B^{-1}X\beta)'WZ)'\end{bmatrix}; \quad A = \begin{bmatrix}
X'X & 0 & ((B^{-1}X\beta)'WX)' \\
0 & \dfrac{n}{\sigma^2} & trWB^{-1} \\
(B^{-1}X\beta)'WX & trWB^{-1} & \sigma^2 trB^{-1}WB^{-1}W+ \\
 & & +[\beta'X'W'WX\beta + \sigma^2 trW'W]
\end{bmatrix}.
$$

References

Abreu, M., De Groot, H., and Florax, R. (2005). Space and Growth: A Survey of Empirical Evidence and Methods. *Région et Développement 21*, 13–44.

Basile, R., Durbán, M., Mínguez, R., Montero, J. M., and Mur, J. (2014). Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control 48*, 229–245.

Billé, A. G., Benedetti, R., and Postiglione, P. (2017). A two-step approach to account for unobserved spatial heterogeneity. *Spatial Economic Analysis 12(4)*, 452–471.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis 28(4)*, 281–298.

Chasco, C., Le Gallo, J., and López, F. A. (2018). A scan test for spatial groupwise heteroscedasticity in cross-sectional models with an application on houses prices in Madrid. *Regional Science and Urban Economics 68*, 226–238.

Chasco C., López F. A., and Le Gallo, J. (2020). The Spatial Structure of Housing Prices in Madrid: Evidence from Spatio-Temporal Scan Statistics. In J. Glaz and M. V. Koutras (Eds.), *Handbook Of Scan Statistics* (1–19). New York: Springer.

Cucala, L. (2016). Scan Statistics for Detecting High-Variance Clusters. *Journal of Probability and Statistics 2016*, Article ID 7591680, 8 pages.

Fischer, M. M., and Stirböck, C. (2006). Pan-European regional income growth and club-convergence. *The Annals of Regional Science 40(4)*, 693–721

Fotheringham, A., Charlton, M., and Brunsdon, C. (1999). Geographically weighted regression. A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A 30(11)*, 1905–1927.

Getis, A., and Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis 24(3)*, 189–206.

Griffith, D. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization.* Berlin, Heidelberg: Springer-Verlag.

Kulldorff, M., and Nagarwalla, N. (1995). Spatial Disease Clusters: Detection and Inference. *Statistics in Medicine 14(8)*, 799–810.

Lauridsen, J., and Kosfeld, R. (2011). Spurious spatial regression and heteroscedasticity. *Journal of Spatial Science 56(1)*, 59–72.

Le Gallo, J., and Dall'Erba, S. (2006). Evaluating the Temporal and Spatial Heterogeneity of the European Convergence Process, 1980–1999. *Journal of Regional Science 46(2)*, 269–288.

Le Gallo, J., López, F. A., and Chasco, C. (2020). Testing for spatial group-wise heteroskedasticity in spatial autocorrelation regression models: Lagrange multiplier scan tests. *The Annals of Regional Science 64(2)*, 287–312.

Lee, J., Gangnon, R. E., and Zhu, J. (2017). Cluster detection of spatial regression coefficients. *Statistics in Medicine 36(7)*, 1118–1133.

López, F. A., Chasco, C., and Le Gallo, J. (2015). Exploring scan methods to test spatial

structure with an application to housing prices in Madrid. *Papers in Regional Science 94(2)*, 317–346.

Mur, J., López, F. A., and Angulo, A. (2008). Symptoms of Instability in Models of Spatial Dependence. *Geographical Analysis 40(2)*, 189–211.

Ramajo, J., Márquez, M. A., Hewings, G. J., and Salinas, M. M. (2008). Spatial heterogeneity and interregional spillovers in the European Union: Do cohesion policies encourage convergence across regions? *European Economic Review 52(3)*, 551–567.

Tango, T., and Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics 4*, 11.

Zhang, Z., Assunção, R., Kulldorff, M. (2010). Spatial Scan Statistics Adjusted for Multiple Clusters. *Journal of Probability and Statistics 2015*, 1–11.