© *Investigaciones Regionales. 21 – Pages 19 to 36*
Section On theory and methods

Investigaciones Regionales

Identifying nonlinear spatial dependence patterns by using non-parametric tests: Evidence for the European Union

Fernando A. López-Hernández *, Andrés Artal-Tur **, M. Luz Maté-Sánchez-Val ***

ABSTRACT: Accounting for spatial structures in econometric studies is becoming an issue of special interest, given the presence of spatial dependence and spatial heterogeneity problems arising in data. Generally, researchers have been employing parametric tests for detecting spatial dependence structures: Moran's I and LM tests in spatial regressions are the most popular approaches employed in literature.However, this approach remains misleading in the presence of nonlinear spatial structures, inducing important biases in the estimation of the parameters of the model. In this paper we illustrate that issue by applying three non-parametrical proposals when testing for spatial structure in data. Empirical findings for the regions of the European Union show important failures of traditional parametric tests if nonlinearities characterise geo-referenced data. Our results clearly recommend employing new families of tests, beyond parametrical ones, when working in such environments.

JEL Classification: C-14, C-63, O-32, R-12.

Keywords: Nonlinear processes, non-parametric tests, spatial dependence, spatial filters, EU regions.

Identificando estructuras espaciales no lineales utilizando test no paramétricos: Evidencias para las Regiones Europeas

RESUMEN: Es cada vez mas frecuente evaluar la presencia de estructuras de dependencia espacial en estudios econométricos cuando se analizan datos de corte

---

* *Corresponding author*: Departamento de Métodos Cuantitativos e Informáticos. Facultad de CC de la Empresa. Universidad Politécnica de Cartagena (UPCT). C/ Real, 3 - 30201 Cartagena (Spain).mailto: *fernando.lopez@upct.es.*

** Departamento de Economía (UPCT).

*** Departamento de Economía Financiera y Contabilidad (UPCT).

Acknowledgments: The authors wishes to thank two anonymous referees for the positive and constructive comments. This work has been carried out with the financial support of project 1897/PHCS/09 of the Fundación Séneca 'Agencia de Ciencia y Tecnología de la Región de Murcia'. F.A. López and M.L. Maté acknowledges financial support from project ECO2009-10534/ECON of the Ministerio de Ciencia y Tecnología of the Reino de España; A.Artal-Tur acknowledges financial support from the FEMISE Association (Project FEM 34-01), the Spanish Ministry of Science and Innovation (ECO 2008-04059/ECON and ECO2011-27169).

Received: 7 april 2011 / Accepted: 12 july 2011.

transversal. La práctica habitual de los investigadores es utilizar tests paramétricos para identificar este tipo de estructuras en los datos y, con diferencia, los dos contrastes más populares son el test de la I de Moran (IM) y el basado en los Multiplicadores de Lagrange (LM). Sin embargo, este enfoque puede ser engañoso cuando en nuestros datos están presentes estructuras de dependencia espacial no lineales. En este trabajo ilustramos esta problemática presentando tres contrastes no paramétricos, alternativos a los clásicos que presentan un mejor comportamiento en presencia de no-linealidades. Una aplicación utilizando diversas variables económicas y filtros espaciales en las Regiones Europeas recomiendan, claramente, utilizar estos contrastes no paramétricos.

Clasificación JEL: C-14, C-63, O-32, R-12.

Palabras clave: Procesos no lineales, contrastes no paramétricos, dependencia espacial, filtros espaciales, Regiones Europeas.

1. Introduction

Spatial models are becoming an important tool in economics, as economists have been rediscovering that geography matters (Anselin, 2010). Research in this area used to begin by applying simple statistics, as Moran's I for example (Moran, 1948), in order to find the presence of a clear spatial pattern in data, and then accounting for it in the subsequent estimation procedure. Nevertheless, traditional parametric statistics, although easy to implement and available in most of the spatial packages, could fail in identifying such correlation patterns in the presence of more complex structures of spatial dependence. As an example, this could be the case when one departs from the linear world, accounting for nonlinear spatial dependence relationships.

Some fields of research have been pioneers in developing nonlinear modeling and accounting for nonlinear relationships in data, given the relevance of obtaining good predictions. Forecasting of exchange rates evolution is one main field in economics where there has been a development of nonlinear methods [1]. The extension of financial crisis is undoubtedly a matter of concern for economists, since the Asian and Latin American turbulences of the 90's, while in recent years has acquired a prominent role fueled by the global financial crisis, which has turned into a sovereign debt crisis. Contagion of financial pressures in the global economy has then become a hot topic in research papers, including nonlinear contagion of financial turbulences leading to national solvency crisis. However, we have just only starting to understand transmission mechanisms of financial and real shocks, and how it affects global financial stability. Extensions of these issues appear of pivotal interest for example for integrated monetary unions as the EU, given that the recent Eurocrisis has unveiled the tough effects that asymmetries between partners can inflict to these areas in case of world financial instability.

Notwithstanding the relevance of the exchange rates topic, the most prominent field of research where we assist to the surge of methodological innovations and de-

[1] See, for example, the early paper of Meese and Rose (1990) on the topic.

partures from the linear world is that one leading with performance of stock markets. Studies focusing in disentangling the presence of nonlinear dependencies in stock returns are becoming very usual in the literature (see, i.e., Hinich and Patterson, 1985). In this field, standard tests of nonlinear dependence have shown strong evidence on the presence of nonlinearities in raw stock returns (Solibakke, 2005). Alternatively, other prominent researchers have been contributing by developing new methods for dealing with nonlinearities in time-series and cross-section modeling, together with neural networks analysis or chaos theory, that have been applied to the study of financial assets behaviour and price formation. As a result, all of these advances have been spilling over the whole profession'smethodological tool-kit, improving our understanding of spatial analysis for socio-economic processes (Lee, White and Granger, 1993).

In this regard,the focus on developing nonlinear models emerges as a clear example of how econometrics is responding to current challenges in data analysis, with new developments arising in the field of spatial econometrics [2]. Some pioneer contributions sharing this focus include those of renamed authors as Arbia *et al.* (2010), Basile (2009), Basile and Girardi (2010) or Osland (2010), that have been showing how non-parametric and semi-parametric techniques can render better results than traditional parametric ones in evaluating nonlinear spatial dependence patterns for cross-sectional data, (López *et al.,* 2010). In summary, employing new proposals better suited for dealing with nonlinearities and the resource to non-parametric and semi-parametric proposals for identifying spatial dependence patterns would surely conform part of the research agenda of spatial econometrics in the near future (Pinkse and Slade, 2010).

In this sense, this paper continues extending that incipient literature: First, we present three types of tests designed for checking for spatial dependence patterns in the presence of nonlinearities: BP test (Brett and Pinske, 1997), Ku test (Kulldorff and Nagarwalla, 1995), and the recently proposed SG test (López et al, 2010). Second, we apply those three proposals on relevant data for the EU regions, as unemployment levels, GDP per capita, etc., in order to empirically capture the emergence of nonlinear spatial structures along that geographical space. And finally, we check for the power of new test against traditional parametric tests (MI) when nonlinearities arise in data analysis. Anticipating some of the results, the failure of the traditional MI test is highlighted in nearly all of the empirical exercises of the investigation. In contrast, non-parametric proposals show greater power in detecting spatial structures in the presence of nonlinearities. In that sense, our results clearly recommend the need of employing new tests in the presence of nonlinearities, given low power of traditional ones.

The remainder of the paper is as follows. In section 2 we make a description of the non-parametric and semi-parametric spatial dependence tests we will apply further in our study. In section 3, we present an empirical application for testing the power of those tests in a nonlinear world. We also include here a discussion of the main findings of the investigation. Finally, section 4 concludes.

---

[2] As an example, consult the monographic number that the *Journal of Econometrics* has recently devoted to the topic (*JoE*, vol. 157 (2010), Elsevier).

2. Non-parametric approach when testing for spatial dependence

In this section, we briefly describe the three non-parametric tests to be employed in the following empirical exercise, commenting on the pros and cons associated to every one of the proposals. We also detail the characteristics of the well-known Moran's I test for novel readers. Further, we evaluate the size dimension of the tests through permutation techniques. All tests are presented in the chronological order they appeared in the literature.

2.1. Four proposals for testing spatial dependence

The most popular test to contrast spatial correlation is Moran's I Test (Moran, 1948) which is widely employed in the first stages of many exploratory and spatial econometrics studies. Moran's I test for a variable x measures if the values of this variable, at different locations (x_i and x_j with $i, j = 1,2, ..., n$ and $i \neq j$), are associated. Formally, Moran's I test follows the expression (1) which is asymptotically distributed as a normal:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \bar{x})w_{ij}(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

where $\bar{x}$ is the sample mean for the variable x, w_{ij} is the (i, j) -element of the known Weight matrix (W) which quantifies the different intensities among spatial locations in function of their proximity. Finally, S_0 is the sum of all W elements and n is the number of observations.

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \tag{2}$$

The second test we present is the Brett and Pinkse proposal (Brett and Pinkse, 1997), that unless appeared more than a decade ago, it is still not so much generalised. This is a non-parametric test which is built considering the properties of the characteristic functions. Specifically, it is based on the property that if two variables (in our case, X and his spatial lag $X^N = WX$) are independent, the joint characteristic function must factorize into the product of their marginal characteristic functions. To compute the test, a f practitioner-chosen density function with infinite support is considered with $h(x) = \int e^{iux} f(u)du$ its Fourier transform. Let $\{X_t\}$ and $\{X_t^N\}$ independent with

X_t^N be the average of proximate observations of X_t. We also define $h_{ts} = h(X_t - X_s)$, $h_{ts}^{NN} = h(X_t^N - X_s^N)$ and, $\eta_{n1} = n^{-2} \sum_{s,t} h_{ts} h_{ts}^{NN}$, $\eta_{n2} = n^{-3} \sum_{s,t,u} h_{ts} h_{tu}^{NN}$ $\eta_{n3} = n^{-4} \sum_{s,t,u,v} h_{ts} h_{uv}^{NN}$, with n the number of observations. Let

$$\eta_n = (\eta_{n1} - \eta_{n2})^2 (\eta_{n2} - \eta_{n3})^2 \tag{3}$$

and

$$\upsilon_n = (\gamma_n - \mu_n^2)^2 n^{-1} \sum_t n_t^{-1} (I(n_t > 0) + \sum_s n_s^{-1} (I(s \in N_t) I(t \in N_s)) \tag{4}$$

where $\mu_n = n^{-2} \sum_{t,s} h_{ts}$, $\gamma_n = n^{-3} \sum_{t,s,u} h_{ts} h_{tu}$, N_t the set of proximate observations of point t and n_t, that is, the cardinal of set N_t.

Then, under the null of independence, the Brett and Pinkse statistic (*BP*)

$$BP = \frac{n\eta_n}{2\upsilon_n} \tag{5}$$

is asymptotically χ_1^2 distributed.

The third alternative is a very popular test in epidemiology (Kulldorff and Nagarwalla, 1995), which has also been employed in economics in its spatial-temporal version (Kang, 2010). In its last definition, due to Kulldorff *et al.* (2009), the Ku test is defined for the case of an underlying Normal distribution, and can be viewed as a semi-parametric test. This proposal, under the null hypothesis, assumes equality of mean values for the variable under study in all locations included in the geo-data set. The alternative hypothesis relies on the existence of a spatial cluster where mean values differ from those of the rest of the sample. In this case where the variable presents spatial structure, and according to the Tobler law, Ku test would reject the null of equidistribution.

Formally, Ku test defines the following null hypothesis,

$$H_0 : X_i \equiv N(\mu,\sigma)\ (\forall i)\ i.i.d.$$

versus the alternative of,

$$H_1 : X_i \equiv N(\mu_Z,\sigma)\ (i \in Z) \text{ and } X_i \equiv N(\lambda_Z,\sigma)\ (i \notin Z) \text{ with } \mu_Z \neq \lambda_Z.$$

where Z is a spatial cluster of connected regions. The new specification of the Ku test allows its generalisation for the analysis of topics related to economics and regional science, widening in that way the scope of research fields where to be applied.

Basically, the Ku test identifies regional clusters where the variable of interest shows significant different behaviour. To define the clusters the test employs «windows» (Z) of different size and shape, then comparing the mean value of the observations lying inside the window with those staying outside it. The «window» (Z) also moves across the entire map, changing its size and shape while searching for identifying the maximum differential existing between the spatial clusters defined in the sample. Once the window with the maximum differential is identified, it is evaluated by checking if that difference appears to be statistically significant. So, under the null hypothesis, the log likelihood of (X_1, ..., X_n) is defined as,

$$\ln L_0 = -n\ln(\sqrt{2\pi}) - n\ln(\sigma) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2} \tag{6}$$

Under the alternative hypothesis, we first calculate the maximum likelihood estimators that are specific to each circle z, which is $\mu_z = x_z/n_z$ with $x_z = \sum_{s\in z} x_s$ and $n_z = \sum_{s\in z} x_n$ for the mean inside the circle, and $\lambda_z = (X - x_z)/(n - n_z)$ with $x_z = \sum_{s\in z} x_s$ for the mean outside the circle. The maximum likelihood estimate for the common variance is,

$$\sigma_z^2 = \frac{1}{n}\left(\sum_{i\in z} x_i^2 - 2x_z\mu_z + n_z\mu_z^2 + \sum_{i\notin z} x_i^2 - 2(X - x_z)\lambda_z + (n - n_z)\lambda_z^2\right) \tag{7}$$

The log likelihood for the alternative hypothesis

$$\ln L_z = -n\ln(\sqrt{2\pi}) - n\ln(\sqrt{\sigma_z^2}) - n/2 \tag{8}$$

Then the Ku statistic is defined as

$$Ku = \max_z(\ln L_z - \ln L_0) = \max_z\left(n\ln(\sigma) + \sum_i \frac{(x_i - \mu)^2}{2\sigma^2} - n/2 - n\ln(\sqrt{\sigma_z^2})\right) \tag{9}$$

Only the last term depends on z, so from this formula it can be seen that the most likely cluster selected is the one that minimizes the variance under the alternative hypothesis, what is intuitive. The p-value is obtained through Monte Carlo hypothesis testing (Dwass, 1957), by comparing the rank of the maximum likelihood from the real data set with the maximum likelihoods from the random data sets. If this rank is r, then the p-value = r/(1 + # simulations). By repeating this procedure and eliminating the selected window we can detect secondary clusters. There is also available a free software to run the Kulldroff test called SatScan, that can be downloaded from *www.satscan.org*.

The final of the proposed tests in our exercise is characterized by a pure non-parametric approach. In comparison with the other two proposals, it does not use the theoretical distribution of observations in its computation. This test, called the SG test, has been proposed recently by one of the authors (López *et al.,* 2010) and builds on the concept of symbolic entropy when defining a measure of cross-sectional spatial dependence. Applying the concept of symbolic entropy for spatial econometrics has been appearing as a feasible tool for dealing with important questions still to be solved in the field (Ruiz *et al.,* 2010; Herrera, 2011).

We explain how to compute the SG test. Given the spatial process $\{X_s\}$with $s \in S$, where S is a set of spatial coordinates, then embedded in an m-dimensional space ($m \geq 2$) as follows:

$$X_m(s_0) = (X_{s_0}, X_{s_1}, \ldots, X_{s_{m-1}}) \tag{10}$$

where s_1, s_2, ..., s_{m-1} are the $m - 1$ closer neighbours to s_0, which are ordered from leaser to greater Euclidean distance with respect to the location s_0. The term $X_m(s)$ is called the m – surroundings of s. The next step in the definition of this test is to encode all the m – surroundings into symbols. To get this purpose, a set of h symbols $\Gamma = \{\sigma_1, \sigma_2, ..., \sigma_h\}$ is defined. Then, the spatial process is symbolised through a symbolization map f with:

$$f : \mathbb{R}^m \to \Gamma \tag{11}$$

such that $f\,[X_m(\mathrm{s})] = \sigma_{j_s}$ with $j_s \in \{1, 2, ..., h\}$. The set of spatial observations $s \in S$ is of σ_i -type if and only if $f\,[X_m(\mathrm{s})] = \sigma_i$.

Based on the symbolization map, the cardinality of the subset S, composed by all the elements of σ_i -type, is defined us $l_{\sigma_i} = \#\{s \in S \,|\, f\,[X_m(s)] = \sigma_i\}$. Besides, the relative frequency of a symbol $\sigma \in \Gamma$ is computed by:

$$p(\sigma) := p_\sigma = \frac{\#\left\{s \in S \mid s \text{ is of } \sigma - \text{type}\right\}}{|S|} \tag{12}$$

where $|S|$ denote the cardinality of the set S.

Under this setting, the symbolic entropy of the spatial process $\{X_s\}$ with $s \in S$ for an embedding dimension is defined as a Shanon's entropy of the h different symbols as follows:

$$q(m) = -\sum_{\sigma \in \Gamma} p_\sigma Ln(p_\sigma) \tag{13}$$

$q(m)$ is the information contained in comparing the m-surroundings generated by the spatial process.

Taking into account previous concepts, the SG test on the spatial process $\{X_s\}$ with $s \in S$ is defined as follows:

$$SG(m) = 2|S|\left[Ln(h) - q(m)\right] \tag{14}$$

This test is asymptotically distributed as a χ_k^2 where k refers to the number of unknown parameters under the alternative hypothesis minus the number of unknown parameters under the null hypothesis.

2.2. Some brief considerations on the characteristics of the spatial tests

In this subsection we briefly review the main features of every defined test, in order to better characterise every one of them. So, the BP test appears to be useful in determining the existence of spatial dependence structures when the underlying spatial process is clearly a nonlinear one (López *et al.,* 2010). In contrast, one of the cons of this test is related to its underlying assumptions, given that the test could fail when the analysed process is a non-stationary one or it does not follow a Normal distribution. So, in the BP-test the spatial process has to be stationary and strongly mixing. In that sense, the BP test requires ex-ante the choice of the function f, with different choices leading to different values of the statistic. In the original paper of Brett and Pinkse, the standard Gaussian density was used for defining the underlying f. In this paper the authors decide to employ the same function according to simulation experiments previously run on a similar time series context by Pinkse (1998), where the author shows any strong sensitivity of the results to the choice of the observations. Also we must note that this aspect of the test has been never explored,neither in its spatial version, nor in its spatio-temporal one (see, i.e, López *et al.,* 2011).

In what respects to the definition of the Kulldorff test, we must note that it does not require any spatial dependence structure information ex-ante, derived from the related weight matrix**.** On the negative side, the Kulldorff test assumes the null hypothesis of *iid*, following a Normal distribution with the same mean value for every cluster or observation in the sample. This is perhaps its more restrictive assumption, with the lack of normality being perhaps responsible in some cases of the rejection of the null of spatial independence. Finally, when implementing the statistic, the researcher must decide the shape of the window and the maximum number of cases that any given window can cover. With the current software available, analysis can be done using circular or elliptical windows (see *www.satscan.com*). The power of the contrast is then related to two factors: (i) The shape of the window Z employed (circular, elliptical or flexible) (ii) The maximum number of elements included in Z. Regarding the first factor, the shape of the window Z used to be defined as a circular window, although employing flexible (computer-defined) shape of windows used to improve the power of the Ku-test (Tango and Takahashi, 2005; Yiannakoulias *et al.,* 2007). In what affects the second factor, it is recommended that the maximum number of cases entering any given window does not exceed 50% of all available cases. In the case that the identified cluster shows a very irregular shape, it is recommended to reduce the number of cases entering the exercise does not surpass 5% or 10% of total available cases. In this paper we follow both recommendations, employing circular windows with a number of cases not accounting for more than 50% of total cases.

In the case of the SG test, its main advantage is related to the fact that it does not require the specification of a pre-determined weight matrix in order to define the neighbouring observations or the underlying spatial structure in data. This is an interesting positive feature of this proposal, but at the same time it does not provide the necessary flexibility to the researcher for testing for the effects of several spatial structures in data. On the negative side, this test renders better results with large than with small samples. Moreover, the SG test present overlapping problems induced by the building of the m-surroundings, which could turn of importance in small samples. Finally, for the SG-test all locations have the same number of neighbours while this not happens for the other two tests.

2.3. Exploring size's tests by employing permutationtechnique

Some properties of the selected tests are described in this subsection, such as the values of the BP test change depending on the chosen *f* function, as well as on the scaling made on observations. The test also presents some problems with the normality assumption if dropped (López *et al.,* 2010). The SG test could also show some size problems in small samples and when using irregular lattices, given overlapping problems. In general, and although it is possible to recover *p-values* from these tests by using asymptotic theory, it seems reasonable to evaluate their behavior by simple permutational test. By doing so, in this subsection we explore the size's characteristics of the proposed test by employing permutation bootstrapping, together with those of the MI test in order to have a reference of a parametrical test. Results on tests' power are not included here for space restrictions, but they are available on request to the authors as usual.

The evaluation of the significance of the coefficients is analysed through the permutation tests. Specifically, for the MI, the Brett and Pinkse and the SG tests a bootstrapping permutation is applied, while for the Kulldorff test a Monte Carlo bootstrapping is undertaken to get the *p-values*. For the BP test we use the proposed transformation suggested by Brett and Pinkse (1997) to drop out scale problems in the variables. To get this purpose, while computing the BP test values, the observations were normalized by first subtracting the median, and subsequently dividing by the median of the absolute values of resulting sequence divided by 0.675 as those authors propose.

Table 1 shows the size values of the considered tests for several sample sizes and distributions. We consider that the observations are distributed on irregular lattices. To compute the MI and BP tests we employ a four nearest neighbour weight matrix. For the SG test, we consider the m-surroundings of size three. The Kuldorff test is built by applying circular windows.

In all cases, independently of the sample lattice or underlying distribution, the sizes appear in the expected range. Therefore, the permutation technique appears

Table 1. Empirical Size. Pseudo *p-value* in irregular lattice

		MI	*BP*	*Ku*	*SG*
N(0,1)	R = 49	0.041	0.055	0.065	0.059
	R = 100	0.054	0.055	0.054	0.051
	R = 225	0.064	0.064	0.064	0.055
U(0,1)	R = 49	0.070	0.054	0.049	0.048
	R = 100	0.060	0.060	0.055	0.067
	R = 225	0.065	0.069	0.065	0.062
$\beta(\frac{1}{2}, \frac{1}{2})$	R = 49	0.050	0.040	0.050	0.040
	R = 100	0.055	0.040	0.055	0.040
	R = 225	0.050	0.060	0.050	0.065
χ_1^2	R = 49	0.060	0,045	0.070	0.070
	R = 100	0.065	0.055	0.045	0.050
	R = 225	0.045	0.050	0.065	0.055

to render good results. Regarding the power of tests, there are some published results that analyze its behaviour in linear and nonlinear processes, using permutation techniques and/or asymptotic theory. Detailed results on the power characteristics for Moran's I, BP, SBDS and SG tests in nonlinear environments can be consulted in López *et al.* (2010), that employ asymptotic theory. A comparison for Moran's I, BP, SG, and Ku tests can be found in López *et al.,* 2011, where the authors employ permutation tests.

3. Nonlinear spatial structures in economic variables: Analysing the case for the European Union

This section undertakes an empirical application to evaluate the behaviour of the previously presented spatial dependence tests under nonlinear process in comparison with the traditional techniques. To get this purpose, we consider as a representative traditional spatial dependence test the MI test of Moran. Because of its simplicity, the Moran's I (MI) test has been widely applied in different research areas. But, the MI test is in strict terms an autocorrelation index, therefore, it not appears as the perfect candidate to evaluate the presence of nonlinear spatial structures in data. The different non-parametric and semi-parametric spatial dependence tests introduced in the previous section could be an interesting alternative to MI for detecting such nonlinear structures in socio-economic variables. The goal of this section is to provide an empirical exercise that illustrates the adequacy of applying alternative non-parametric or semi-parametric spatial dependence tests when we presume the existence of nonlinear spatial dependence structures in the data.

In our empirical exercise, we will use both the Cambridge Econometrics and REGIO databanks. From these databases, we focus our analysis on a total of 261 regions, NUTS II level, from the 27 countries that are currently members of the European Union (EU-27). For different reasons, various regions have been excluded, among them the Canary Islands, Ceuta, Melilla and the Portuguese archipelagos of the Azores and Madeira. With the aim of providing more robustness to our results, we develop our analysis for three years (1991, 2000 and 2010). We focus our attention on four classic variables computed for the European Regions because of their importance as economic indicators. These are: the unemployment rate (UR), the percentage of active population in the agriculture sector over the total population (EAr), the R & D expenditure per capita (RDpc) and the gross domestic product per capita (GDPpc).

3.1. Spatial dependence structure in the original raw data

Figure 1 shows the Box Plot of the analysed variables for the last year of the sample, 2010. In all cases we observe a clear spatial dependence structure: For ex-

Figure 1. Quartile Map for original variables (year 2010)

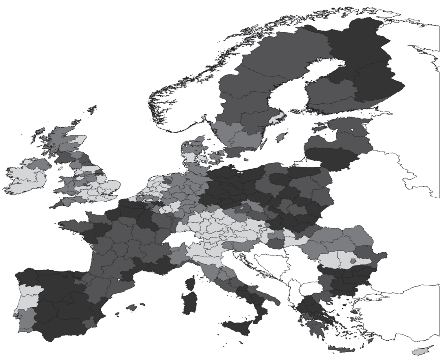

Figure 1a. Unemployment rate

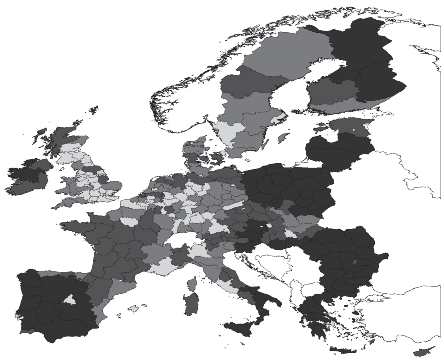

Figure 1b. Agricultural employment rate

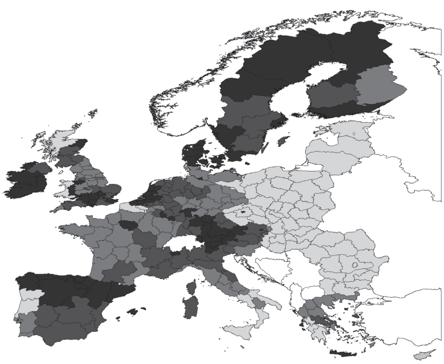

Figure 1c. R&Dpc

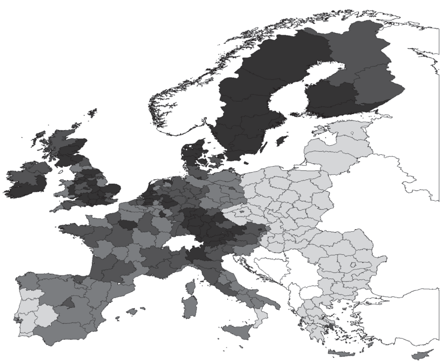

Figure 1d. GDPpc

ample, for the unemployment rate (UR) variable, the highest values correspond to the periphery regions of the Eastern Europe, together with some southern EU regions in Spain, Greece, Italy, of the former Yugoslavia. Agricultural employment (AEr) rate is also higher than the EU average in southern and eastern regions, showing the structural socio-economic changes that these territories are still facing. In terms of R&D expenditures and GDP per capita, the figure shows the contrary situation, with regions in Scandinavian countries (Finland, Denmark and Sweden, particularly) showing the highest rate of investments and living standards or purchasing power. Other European regions in Germany, United Kingdom or France (Ille-de-France) also occupy an important position in that ranking, showing a clear spatial dependence structure along the EU space for all of the chosen variables.

Table 2 shows the values for the different spatial dependence tests and their pseudo p-values. Again for all cases, the statistical values appear very high, leading to a rejection of the null hypothesis about a random pattern in the spatial distribution of data.

Table 2. Test of Diagnostic for spatial dependence on the original data (y)

	MI	*p-value*	*SG*	*p-value*	*BP*	*p-value*	*Ku*	*p-value*
UR 1991	11.12	0.000	188.3	0.000	707.2	0.000	19.6	0.004
UR 2000	12.16	0.000	202.4	0.000	868.0	0.000	22.1	0.003
UR 2010	15.29	0.000	159.5	0.000	941.8	0.000	46.4	0.000
AEr 1991	16.57	0.000	107.8	0.000	1,089.5	0.000	89.5	0.000
AEr 2000	17.43	0.000	129.7	0.000	832.9	0.000	99.6	0.000
AEr 2010	17.03	0.000	94.9	0.000	736.9	0.000	94.7	0.000
RDpc 1991	17.45	0.000	292.8	0.000	3,270.9	0.000	81.9	0.000
RDpc 2000	16.71	0.000	226.9	0.000	2,266.4	0.000	86.6	0.000
RDpc 2010	14.62	0.000	181.1	0.000	1,546.8	0.000	66.6	0.000
GDPpc 1991	18.22	0.000	289.5	0.000	5,400.9	0.000	90.9	0.000
GDPpc 2000	17.38	0.000	213.2	0.000	4,509.2	0.000	93.2	0.000
GDPpc 2010	16.25	0.000	187.7	0.000	3,307.8	0.000	76.1	0.000

p-value = p-seudo value obtain test by permutational bootstrapping. 999 iterations.

The next step in our empirical application is now dropping from these variables the linear spatial dependence structure. In order to do so, we apply the filtering technique usually employed in the spatial econometrics literature, namely the Getis (1990, 1995) proposal [3].

[3] Equivalent results are obtained by authors when filtering data using simple spatial autoregressive model estimation. To use this technique we estimate a simple spatial autoregressive model for each of the empirical variables. Therefore, the residuals of this estimation should not contain any spatial dependence structure. Results are available upon request as usually.

3.2. Applying the filter of Getis

Among the most commonly applied spatial filtering techniques we find the Getis (1990, 1995) proposal, as well as the Griffith (1996, 2003) eigenvector spatial filtering approach. A recent empirical comparison of that two filtering techniques, spatial lag regression and Getis filtred, has shown that both approaches are almost equally equipped for removing the spatial effects from geographically organized variables (Getis and Griffith, 2002). Given their similar empirical performance, for the remainder of the paper we rely on the Getis approach, which has been applied in a variety of empirical research contexts (see e.g. Badinger *et al.,* 2004; Battisti and Di Vaio, 2008; Mayor and López, 2008). Moreover, as Getis and Griffith (2002) argue, the advantage of the Getis approach compared to the eigenvector filtering relies in its simplicity.

To derive the set of spatially «cleaned» variables, the Getis approach uses the local statistic $G_i(d)$ (Getis and Ord, 1992). So, the new filtered variable is defined as in (5)

$$y_i{**} = \frac{y_i\left(W_i/R-1\right)}{G_i(d)} \tag{15}$$

where $G_i(d)$ is the local statistic of Getis and Ord and W_i is the sum of the i- row of the contiguity W matrix. The transformation procedure depends on identifying an appropriate distance d within which nearby areal units are spatially dependent. There have been suggestions for identifying this magnitude d. One of which requires that the statistic $G_i(d)$ be evaluated at a series of increasing distances until no further spatial autocorrelation is evident.

With the aim of filtering data we choose a weight matrix based on the Euclidean distance. Nevertheless, some of the European regions in our sample are located at a long distance from the others. This regional disposition breaks with the symmetry in the weight matrix and needed for computing the local index. To overcome this situation, we connect the furthest regions with the two closers locations independently of the Euclidean distance.

$$w_{ij} = \begin{cases} 1 \text{ if } \; d_{ij} < d \text{ or } j \in NN(i,2) \\ 0 \qquad\qquad\qquad \text{in other case} \end{cases} \tag{16}$$

where NN(i,2) is the set of two nearest neighbors to «i». Figure 2 shows the Box Plots for the filtered variables (y**) by applying the previously described procedure. In this case, results seem to be clearer than for the previous analysis, with the other filtering technique: There is not graphical evidence about the existence of spatial dependence structures.

Figure 2. Quartile Map for Getis Filter variable (y**) (year 2010)

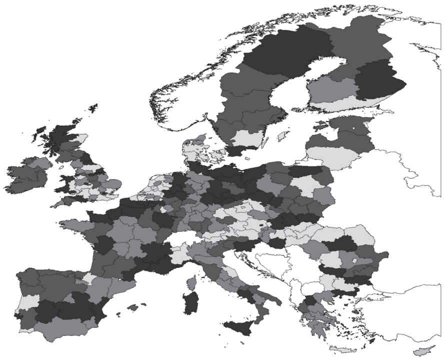

Figure 2a. Unemployment rate

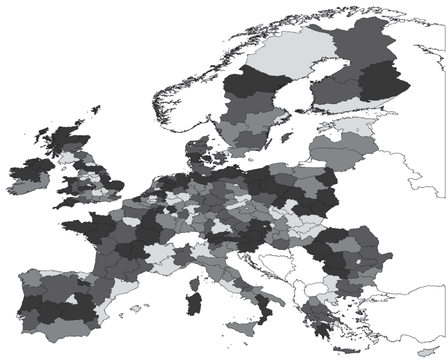

Figure 2b. Agricultural employment rate

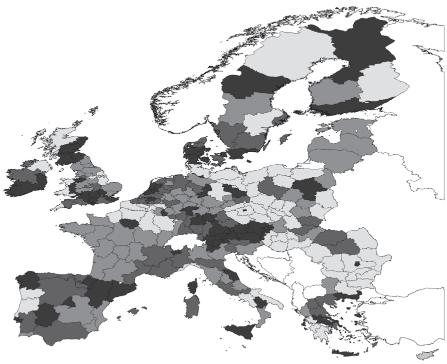

Figure 2c. R&Dpc

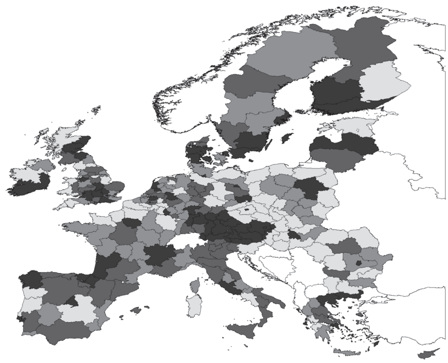

Figure 2d. GDPpc

Table 3 presents the statistical results for our four spatial dependence tests on the filtered variable (y**). For each variable, the shorter distance (*d* in kilometers) is selected in order to drop the spatial dependence structure according to the Getis filter.

Table 3. Test of diagnostic spatial dependence on Getis filtered variables (y**)

	d	*MI*	*p-value*	*SG*	*p-value*	*BP*	*p-value*	*Ku*	*p-value*
UR 1991	280	–1.39	0.150	3.2	0.510	35.3	0.002	9.0	0.604
UR 2000	260	–1.60	0.091	2.5	0.593	40.4	0.000	7.4	0.995
UR 2010	260	–1.11	0.248	2.6	0.557	26.1	0.004	7.4	0.966
AEr 1991	480	–1.17	0.207	3.7	0.390	0.6	0.520	8.1	0.319
AEr 2000	420	–1.04	0.292	6.6	0.167	0.3	0.691	12.5	0.153
AEr 2010	420	–1.35	0.144	7.3	0.135	1.3	0.303	6.5	0.594
RDpc 1991	340	–1.00	0.301	8.5	0.090	25.1	0.007	19.4	0.036

Table 3. *(Continue)*

	d	*MI*	*p-value*	*SG*	*p-value*	*BP*	*p-value*	*Ku*	*p-value*
RDpc 2000	400	–1.11	0.253	32.5	0.000	39.4	0.002	13.0	0.235
RDpc 2010	380	–1.34	0.151	19.2	0.006	2.8	0.242	10.8	0.213
GDPpc 1991	360	–0.73	0.432	25.0	0.001	14.5	0.024	32.1	0.024
GDPpc 2000	380	–1.20	0.210	32.4	0.000	8.9	0.048	23.8	0.048
GDPpc 2010	400	–1.16	0.214	35.3	0.000	7.2	0.079	13.0	0.240

d= distance in Km to compute W(d). p-value = pseudo p-value obtained by permutational bootstrapping (999 iterations).

According to the results of the Moran test, no one of the selected variables would present further spatial dependence signs. On the other hand, the non-parametric tests allows us to observe still the presence of spatial structures in data, with pseudo *p-values* higher than 0.05, what lead to the rejection of the null hypothesis of independence. In that way, for this filtering technique we observe similar conclusions than those obtained after applying the spatial lag filter, once we apply the non-parametric or semi-parametric contrasts. In summary, the resource to such new proposals has allowed us to unequivocally detect the spatial dependence structure underlying our socio-economic variables from a nonlinear perspective. The behavior of the non-parametric and semi-parametric tests in comparison to the traditional spatial dependence tests (Moran's I) highlights the relevance of their application in the initial steps of every spatial dependence analysis with traces of nonlinear spatial structures. The absence of this battery of tests in the researcher's tool kit could obviously generate negative effects in her/his posterior econometric estimation process (Le Sage and Pace, 2009), as we have been able to show in this paper.

Analyzing the results for each variable, we get that for the Agricultural Employment rate (AEr) the spatial dependence structure is completely dropped through the Getis filtering technique. In this sense, all spatial dependence tests accept the null hypothesis of independence. This result is not similar for the other studied variables, particularly, in the case of the RDpc and GDPpc variables where tests reject the null of independence for the years 1991 and 2000. In these cases, the proposed non-parametric and semi-parametric tests are able to capture the presence of spatial structures in a nonlinear fashion. A similar conclusion is found for the Unemployment rate (UR) when the BP test is employed. All of these render important conclusions for the spatial econometrics literature, particularly in the presence of nonlinearities.

4. Conclusions

The interaction relationships among spatial units are complex in empirics. Identifying those linkages is not always a simple matter and, because of that, specifying spatial structures through linear models is not always the best modeling option. The fact that some tests, for example the Moran's I test, have become popular among researchers because of its simplicity and the availability of friendly software to run the computing process, should be complemented with other alternative tests, given the low power characterizing that simple spatial correlation test. Therefore, there is a need in the literature of spreading knowledge on alternative tools useful for evaluating the presence of spatial dependence structures in geo-data.

In this paper, we have tested the improvements that several non-parametric tests can provide to empirical analysis when nonlinear dependence structures could be present in data, this being the pivotal contribution of the investigation. This is an important point, given that some renamed authors as Anselin and Florax (1995) insist in what MI test is a general specification contrast, although they do not really address its weakness in a nonlinear world. Given that Moran's I could fail in detecting spatial association when we depart from simple dependence structures, as we have shown along the empirical part of the paper, we have proposed to employ three new tests recently developed, namely Kulldorff, BP and SG tests. All of them have shown to be well endowed for detecting spatial structures in the presence of nonlinearities. However, we have also shown that everyone performs better under particular circumstances, depending on the distributional characteristics of the process to be analyzed.

In summary, our investigation has shown the importance of following new proposals when testing for spatial correlation if one wants to depart from the linear world. On the contrary, results of econometric modeling could induce important biases when estimating parametersof interest, taking to potential misleading results in policy terms, and a waste of scarce public funds, something very important in a period of hard budgetary constrains as this is.

5. References

Anselin, L. (2010): «30 years of Spatial Econometrics», *Papers in Regional Science*, 89 (1): 3-25.

Anselin, L., and Florax, R. (1995): Small sample properties of tests for spatial dependence in regression models: Some further results. In: Anselin, L., Florax, R. (eds) *New directions in spatial econometrics*. Springer, Berlin.

Arbia, G.; Espab, G.; Giulianic, D., and Mazzitellic, A. (2010): «Detecting the existence of space-time clustering of firms», *Regional Science and Urban Economics*, 40 (5): 311-323.

Badinger, H.; Müller, W., and Tondl, G. (2004): «Regional Convergence in the European Union, 1985-1999: A Spatial Dynamic Panel Analysis», *Regional Studies*, 38: 241-253.

Basile, R. (2009): «Regional economic growth in Europe: A semiparametric spatial dependence approach», *Papers in Regional Science,* 87 (4): 527-565.

Basile, R., and Girardi, A. (2010): «Specialization and risk sharing in European regions», *Journal of Economic Geography*, 10 (5): 645-659.

Battisti, M., and Di Vaio, G. (2008): «A spatially filtered mixture of convergence regressions for EU regions, 1980-2002», *Empirical Economics,* 34: 105-121.

Brett, C., and Pinkse, J. (1997): «Those taxes are all over the map! A test for spatial independence of municipal tax rates in British Columbia», *International Regional Science Review*, 20: 131-151.

Dwass, M. (1957): «Modified randomization tests for nonparametric hypotheses», *Annals of Mathematical Statistics*, 2: 181-187.

Getis, A. (1990): «Screening for Spatial Dependence in Regression Analysis», *Papers of the Regional Science Association*, 69: 69-81.

— (1995): «Spatial Filtering in a Regression Framework: Experiments on Regional Inequality, Government Expenditures, and Urban Crime», in Anselin, L., Florax, R. and Rey, S. (eds.): *New Directions in Spatial Econometrics*, Berlin, Springer.

— and Ord, J. K. (1992): «The analysis of spatial association by use of distance statistics», *Journal of Geographical Analysis*, 24: 189-206.

— and (1995): «Local spatial autocorrelation statistics: Distributional issues and an application», *Journal of Geographical Analysis*, 27: 286-305.

— and Griffith, D. A. (2002): «Comparative spatial filtering in regression analysis», *Geographical Analysis*, 34 (2): 130-140.

Griffith, D. A. (1996): «Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data», *The Canadian Geographer*, 40: 351-367.

— (2003): *Spatial Autocorrelation and Spatial Filtering*, Berlin, Springer.

Herrera, M. (2011): «Causality. Contributions to Spatial Econometrics», Ph. D. Thesis.University of Zaragoza.

Hinich, M. J., and Patterson, D. M. (1985): «Evidence of Nonlinearity in Daily Stock Returns», *Journal of Business & Economic Statistics*, 3 (1): 69-77.

Kang, H. (2010): «Detecting agglomeration processes using space-time clustering analyses», *Annals of Regional Science,* 45: 291-311.

Kulldorff, M., and Nagarwalla, N. (1995): «Spatial disease clusters: Detection and Inference», *Statistics in Medicine*, 14: 799-810.

Kulldorff, M.; Huang, L., and Konty, K. (2009): «A scan statistic for continuous data based on the normal probability model», *International Journal of Health Geographics*, 8, 58-73.

Le Sage, J., and Pace, R. K. (2009): *Introduction to Spatial Econometrics,* England, Taylor and Francis Group.

Lee, T. H.; White, H., and Granger, C. W. J. (1993): «Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests», Journal of Econometrics, 56 (3): 269-290.

López, F. A.; Maté, M. L., and Artal, A. (2011): Evaluating three proposals for testing independence in non linear spatial processes.Fundación de las Cajas de Ahorros. Working paper, n.º 622/2011.

López, F. A.; Matilla, M.; Mur, J., and Ruiz, M. (2010): «A non-parametric spatial independence test using symbolic entropy», *Regional Science and Urban Economics* 40, 106-115.

— (2011): «Four tests of independence in spatio-temporal data», *Papers in Regional Science.* 90 (3): 663-685.

Mayor, M., and López, A. (2008): «Spatial shift-share analysis versus spatial filtering: an application to Spanish employment data», *Empirical Economics*, 34: 123-142.

Meese, R. A., and Rose, A. K. (1990): «Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation», *The American Economic Review*, 80 (2), Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association: 192-196.

Moran, P. A. (1948): «The interpretation of statistical maps», *Journal of the Royal Statistical Society*, Series B(10): 243-51.

Osland, L. (2010): «An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling», *Journal of Real Estate Research,* 32 (3): 289-320.

Pinkse J., and Slade, M. E. (2010): «The future of spatial econometrics», *Journal of Regional Science*, 50: 103-117.

Pinkse, J. (1998): «A consistent non-parametric test for serial independence», *Journal of Econometrics*, 84 (2): 205-231.

Ruiz, M.; López, F. A., and Páez, A. (2010): «Testing for spatial association of qualitative data using symbolic dynamics», *Journal of Geographical Systems*, 12 (3): 281-309.

Solibakke, P. (2005): «Non-linear dependence and conditional heteroscedasticity in stock returns evidence from the norwegian thinly traded equity market», *European Journal of Finance*, 11 (2): 111-136.

Tango, T., and Takahashi, K. (2005): «A flexibly shaped spatial scan statistic for detecting clusters», *International Journal of Health Geographics*, 4 (1): 11.

Yiannakoulias, N.; Rosychuk, R., and Hodgson, J. (2007): «Adaptations for finding irregularly shaped disease clusters», *International Journal of Health Geographics,* 6 (1): 28.