A Non-Parametric Spatial Independence Test Using Symbolic Entropy

Ruiz Marín, Manuel (manuel.ruiz@upct.es)
Dpto. Métodos Cuantitativos e Informáticos, Universidad Politécnica de Cartagena
Paseo Alfonso XIII, 50 - 30.203. Cartagena. Spain
López Hernández, Fernando (fernando.lopez@upct.es)
Dpto. Métodos Cuantitativos e Informáticos, Universidad Politécnica de Cartagena
Paseo Alfonso XIII, 50 -30.203. Cartagena. Spain

RESUMEN

In the present paper, we construct a new, simple, consistent and powerful test for spatial independence, called the SG test, by using symbolic dynamics and symbolic entropy as a measure of spatial dependence. We also give a standard asymptotic distribution of an affine transformation of the symbolic entropy under the null hypothesis of independence in the spatial process. The test statistic and its standard limit distribution, with the proposed symbolization, are invariant to any monotonuous transformation of the data. The test applies to discrete or continuous distributions. Given that the test is based on entropy measures, it avoids smoothed nonparametric estimation. We include a Monte Carlo study of our test, together with the well-known Moran's I, the $SBDS$ (de Graaff et al, 2001) and τ (Brett and Pinkse, 1997) non parametric test, in order to illustrate our approach.

Palabras clave: Spatial independence,non-parametric, entropy.

Clasificacin JEL (Journal Economic Literature): C21; C50; R15.

rea temtica: Mtodos Cuantitativos

1 Introduction

Dependence is one of the most outstanding characteristics of spatial data. Gould (1970, pp 443-444) asks '*Why we should expect independence in spatial observations*' and his answer is simple: '*I cannot imagine*'. Tobler (1970, p.237) goes a step further when he refers to '*the first law of geography: everything is related to everything else*'. Along the same lines, Anselin (1988, p.12) proclaims, '*The essence of regional science (...) is that location and distance matter, and result in a variety of interdependencies in space-time*' and Paelinck and Klaassen (1979, p.5) state that, '*(...) it is good to start out in spatial econometric modelling with an interdependent model*'. To sum up, there is a strong consensus about the importance of this question (Getis 2007), which already forms a routine part of any spatial Econometric application.

The first problem in this discussion is to detect when the hypothesis of independence is not admissible, for which it will be necessary to use some of the tests proposed in the literature. There is a wide variety of proposals and we could distinguish up to five categories. (1) The traditional approach based on the space-time interaction coefficient of Knox (1964), of which we can consider the tests of Moran (1950), Geary (1954) and Dacey (1965), among others, as particular cases. (2) Anselin (1988)'s text fully introduces the maximum-likelihood methodology into this field, along with a new generation of more specific and flexible tests (Anselin et al, 1996, Anselin and Bera 1988, Anselin, 2001, Leung et al, 2003, and Baltagi et al, 2003). (3) Kelejian and Robinson (1993) propose using instrumental variables in connection with spatial models, which leads to a new battery of tests of spatial dependence based, directly or indirectly, on the GMM principle (Anselin and Kelejian, 1997, Kelejian and Prucha, 1999, Conley, 1999, Saavedra, 2003, Kelejian and Prucha, 2004 and 2007, and Fingleton, 2008). (4) The KR test of Kelejian and Robinson (1992) and the Lagrange Multiplier, which Anselin and Moreno (2003) derive for the error components model of Kelejian and Robinson (1995). (5) The last type of tests incorporated into the analysis of spatial data are the non-parametric tests like the $SBDS$ test (de Graaff et al, 2001) and the τ test (Brett and Pinkse, 1997, Pinkse et al, 2002).

In this paper, we propose a new test, called SG, whose immediate predecessor is the test of serial independence in a time series developed by Matilla and Ruiz (2008) based on permutation entropy. In general, the measures associated with entropy, applied in a context of time series, have gained importance in recent years (see Joe, 1989a and b, Hong and White, 2005, and references therein) although, as far as we know, this is the first time that the approach has been used in a context of spatial data. It is a non-parametric test, not very demanding in terms of a priori hypotheses. Furthermore, with the symbolization proposed, it is consistent and invariant to any monotonous transformation of the series and its asymptotic distribution function is standard. If we add that it is easy to obtain and that it is competitive against other well-established tests in the literature, we think that it could play an interesting role in the toolbox of spatial data analysis. The peculiarity of the SG test is that it uses symbolic entropy as a measure of cross-sectional dependence. The idea is simple and is carried out in two stages (Matilla and Ruiz, 2008, p.2-3). The first is to symbolize the series through '*symbol sequences obtained for a suitable partition of the state space*' whose mission is to capture the dynamic structure in the series. Then, the result is interpreted, in the light of the Theory of Information: '*we use the entropy measure associated to these symbols to test the dependence present in the (...) series*'.

In order to better appreciate the characteristics of the SG test, with respect to the others found in the literature, we are going to refer to the following aspects:

- Dependence vs. spatial autocorrelation.

- Linearity vs. nonlinearity.

- Normality.
- The role of the weighting matrix.

The tests to which we have referred at the beginning of this section are mostly non autocorrelation tests. For example, in the test of Moran, we test whether the covariance between the series and its spatial lag is statistically different from zero. Moreover, the maximum-likelihood tests are linked directly to a coefficient of autocorrelation. Nevertheless, non autocorrelation is synonymous with independence only under restrictive conditions (as in gaussian stationary random fields; see Arbia, 1989 and 2006, for a deeper discussion of the concept of spatial random field). The SG test is, in a strict sense, a test of independence like the $SBDS$ and τ tests, although it is more generic than them (in the $SBDS$ test, the structure of dependencies must be absolutely regular, while that of Brett and Pinkse requires strongly mixing processes). The assumption of linearity is not an essential requirement in non-parametric tests. This is an important characteristic because the tests that use the linear correlation approach are not consistent against other alternatives of non-linear dependence with zero autocorrelation like the non-linear moving average processes or the spatial ARCH (called SARCH processes by Bera and Simlai, 2004). Normality is a minor restriction but forms part of the set of hypotheses on which the tests habitually used in a spatial context are based. The assumption is of the utmost importance for the maximum-likelihood tests and is very useful in those linked to the Knox statistic. The exact distribution of Moran's test, even assuming normality is not standard and depends on the eigenvalues of the weighting matrix (Tiefelsdorf and Boots, 1995, Tiefelsdorf, 2000, Kelejian and Prucha, 2001). Under relatively weak conditions, this distribution converges to the normal distribution (Sen, 1976). If the assumption of normality is not acceptable, and to avoid biases in the test, Cliff and Ord (1981) propose using a type of bootstrap that they call '*randomisation*'. This discussion, in the case of the three non-parametric tests ($SBDS$, τ and SG), is more simple. Their exact distribution is unknown although, asymptotically and whatever happens with the finite sampling distribution, we obtain standard distributions (a normal one in the case of $SBDS$ and chi-squared with the τ and SG tests).

The absence of a natural ordering of the data is an inevitable source of problems when dealing with spatial series. The usual solution is to specify a weighting matrix using '*a set of weights which he (the investigator) deems appropriate from prior considerations*' (Cliff and Ord, 1981, p.17). This situation is very undesirable because it implies that the test not only examines the existence or not of spatial dependence in the data, but also the adequacy of the weighting matrix itself. In fact, as Pinskse (2004) indicates, this matrix forms part of the null hypothesis. Florax and Rey (1995) demonstrate that, if the matrix is misspecified, the tests tend to lose reliability. The consequences will be more severe, the more serious the misspecification (Cliff and Ord, 1981, p.168, with respect to Moran's I). The key term in this case is '*uncertainty*', although some authors prefer to speak of '*flexibility*', and it remains to be one of the fundamental problems in applied spatial econometric modeling (Griffith, 1996, Bavaud, 1997, Haining, 2003, for a discussion). In this context, we wish to underline that the SG test, the same as the $SBDS$, does not require the specification of a weighting matrix, unlike the other tests (in the τ test it is necessary to specify the neighbors of each point, which is equivalent to constructing the whole weighting matrix).

The paper consists of seven sections. In the second, we introduce the concepts and the basic notation that we will use in the rest of the paper. In the third section, we construct the test of independence, based on symbolic entropy, that motivates our research. The fourth section discusses the most important properties of the test. The fifth section is dedicated to the symbolization procedure of the series, with respect to which the user has a lot of flexibility. The sixth section presents the results of a Monte Carlo experiment in which we examine the behavior of the SG test together with Moran's I, the $SBDS$ and τ tests. The paper finishes with a section of conclusions and future perspectives.

2 Definitions and Notation

In this section we give some definitions and we introduce the basic notation.

Let $\{X_s\}_{s\in S}$ be a real-valued spatial process, where S is a set of coordinates. Given a coordinate s_0 we will denote by (ρ_i^0, θ_i^0) the polar coordinates of location s_i taking as a origin s_0.

Let $\Gamma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ be a set of n symbols. Let m be a natural number with $m \geq 2$. Next, we consider that the spatial process $\{X_s\}_{s\in S}$ is embedded in an m-dimensional space as follows:

$$X_m(s_0) = (X_{s_0}, X_{s_1}, \ldots, X_{s_{m-1}}) \text{ for } s_0 \in S \tag{1}$$

where $s_1, s_2, \ldots, s_{m-1}$ are the $m-1$ nearest neighbors to s_0 satisfying the following two conditions:

(a) $\rho_1^0 \leq \rho_2^0 \leq \ldots \leq \rho_{m-1}^0$,

(b) and if $\rho_i^0 = \rho_{i+1}^0$ then $\theta_i^0 < \theta_{i+1}^0$.

Notice that conditions (a) and (b) ensure the uniqueness of $X_m(s)$ for all $s \in S$. We will call $X_m(s)$ an m-surrounding of s.

Now assume that there is a map

$$f : R^m \to \Gamma \tag{2}$$

defined by $f(X_m(s)) = \sigma_{j_s}$ with $j_s \in \{1, 2, \ldots, n\}$. We will say that $s \in S$ is of σ_i-type if and only if $f(X_m(s)) = \sigma_i$. We will call the map f a *symbolization map.*

Moreover, if the symbolization map f is such that, under the null of independence, all the symbols have the same probability of occurring, we will say that f is a *standard symbolization map*.

Denote by

$$n_{\sigma_i} = \sharp\{s \in S|\ f(X_m(s)) = \sigma_i\}, \tag{3}$$

that is, the cardinality of the subset of S formed by all the elements of σ_i-type.

Also, under the conditions above, one could easily compute the relative frequency of a symbol $\sigma \in \Gamma$ by:

$$p(\sigma) := p_\sigma = \frac{\sharp\ \{s \in S \mid s \text{ is of } \sigma - \text{type}\}}{|S|} \tag{4}$$

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

Now, under this setting, we can define the *symbolic entropy* of a spatial process $\{X_s\}_{s\in S}$ for an embedding dimension $m \geq 2$. This entropy is defined as the Shanon's entropy of the n distinct symbols as follows:

$$h(m) = -\sum_{\sigma\in\Gamma} p_\sigma \ln(p_\sigma). \tag{5}$$

Symbolic entropy, $h(m)$, is the information contained in comparing the m-surroundings generated by the spatial process. Notice that, in the case in which the symbolization map is standard, $0 \leq h(m) \leq \ln(n)$ where the lower bound is attained when only one symbol occurs, and the upper bound for a completely random system (i.i.d. spatial sequence) where all n possible symbols appear with the same probability.

3 Construction of the Independence Test

In this section, we construct an independence test with all the machinery defined in Section 2. We also prove that an affine transformation of the symbolic entropy defined in (5) is asymptotically χ^2 distributed.

Let $\{X_s\}_{s\in S}$ be a spatial process and m be a fixed embedding dimension. In order to construct a test for spatial independence in $\{X_s\}_{s\in S}$, which is the aim of this paper, we consider the following null hypothesis:

$$H_0 : \{X_s\}_{s\in S} \quad i.i.d \tag{6}$$

against any other alternative.

Now, for a symbol $\sigma_i \in \mathcal{S}$, we define the random variable $Z_{\sigma_i s}$ as follows:

$$Z_{\sigma_i s} = \begin{cases} 1 & \text{if } f(X_m(s)) = \sigma_i \\ \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

that is, we have that $Z_{\sigma_i s} = 1$ if and only if s is of σ_i-type, $Z_{\sigma_i s} = 0$ otherwise.

Then $Z_{\sigma_i s}$ is a Bernoulli variable with probability of "success" p_{σ_i}, where "success" means that s is of σ_i-type. It is straightforward to see that

$$\sum_{i=1}^{n} p_{\sigma_i} = 1 \tag{8}$$

Now assume that set S is finite and of order R. Then we are interested in knowing how many s's are of σ_i-type for all symbol $\sigma_i \in \mathcal{S}$. In order to answer the question, we construct the following variable

$$Y_{\sigma_i} = \sum_{s=1}^{T} Z_{\sigma_i s} \tag{9}$$

The variable Y_{σ_i} can take the values $\{0, 1, 2, \ldots, R\}$. Then it follows that the variable Y_{σ_i} is the Binomial random variable

$$Y_{\sigma_i} \approx B(R, p_{\sigma_i}). \tag{10}$$

Then under the null H_0, the joint probability density function of the n variables $(Y_{\sigma_1}, Y_{\sigma_2}, \ldots, Y_{\sigma_n})$ is:

$$P(Y_{\sigma_1} = a_1, Y_{\sigma_2} = a_2, \ldots, Y_{\sigma_n} = a_n) = \frac{(a_1 + a_2 + \ldots + a_n)!}{a_1!a_2! \cdot \ldots \cdot a_n!} p_{\sigma_1}^{a_1} p_{\sigma_2}^{a_2} \cdots p_{\sigma_n}^{a_n} \tag{11}$$

where $a_1 + a_2 + \ldots + a_n = R$. Consequently, the joint distribution of the n variables $(Y_{\sigma_1}, Y_{\sigma_2}, \ldots, Y_{\sigma_n})$ is a multinomial distribution.

Then the likelihood ratio statistic is (see, for example, Lehmann, 1986):

$$\begin{aligned}\lambda(Y) &= \frac{\frac{R!}{n_{\sigma_1}!n_{\sigma_2}!\cdot\ldots\cdot n_{\sigma_n}!} p_{\sigma_1}^{n_{\sigma_1}} p_{\sigma_2}^{n_{\sigma_2}} \cdots p_{\sigma_n}^{n_{\sigma_n}}}{\frac{R}{n_{\sigma_1}!n_{\sigma_2}!\cdot\ldots\cdot n_{\sigma_n}!} \widehat{p}_{\sigma_1}^{n_{\sigma_1}} \widehat{p}_{\sigma_2}^{n_{\sigma_2}} \cdots \widehat{p}_{\sigma_n}^{n_{\sigma_n}}} = \frac{\prod_{i=1}^{n} p_{\sigma_i}^{n_{\sigma_i}}}{\prod_{i=1}^{n} \left(\frac{n_{\sigma_i}}{R}\right)^{n_{\sigma_i}}} = \\ &= R^{\sum_{i=1}^{n} n_{\sigma_i}} \prod_{i=1}^{n} \left(\frac{p_{\sigma_i}}{n_{\sigma_i}}\right)^{n_{\sigma_i}} = R^R \prod_{i=1}^{n} \left(\frac{p_{\sigma_i}}{n_{\sigma_i}}\right)^{n_{\sigma_i}}. \end{aligned} \tag{12}$$

On the other hand, $SG(m) = -2\ln(\lambda(Y))$ asymptotically follows a Chi-squared distribution with k degrees of freedom, where k is equal to the number of unknown parameters under H_1 minus the number of unknown parameters under H_0 (see, for instance, Lehmann, 1986).

Now, if symbolization map f is standard, that is, under the null hypothesis, all the symbols have the same probability of occurring, $p_{\sigma_i} = \frac{1}{n}$ for all $i = 1, 2, \ldots, n$, then it follows that

$$SG(m) = -2R[\ln(\frac{1}{n}) + h(m)] = -2R[h(m) - \ln(n)] = 2R[\ln(n) - h(m)]. \tag{13}$$

Therefore, we have proved the following theorem.

Theorem. *Let $\{X_s\}_{s\in S}$ be a real-valued spatial process with $|S| = R$. Assume that there exist a standard symbolization map f for $\{X_s\}_{s\in S}$. Denote by $h(m)$ the symbolic entropy defined in (5) for a fixed embedding dimension $m \geq 2$. If the spatial process $\{X_s\}_{s\in S}$ is i.i.d., then the affine transformation of the symbolic entropy*

$$SG(m) = 2R[\ln(n) - h(m)] \tag{14}$$

is asymptotically χ^2_k distributed.

Let α be a real number with $0 \leq \alpha \leq 1$. Let χ^2_α be such that

$$P(\chi^2_k > \chi^2_\alpha) = \alpha. \tag{15}$$

Then to test

$$H_0 : \{X_s\}_{s \in S} \quad i.i.d. \tag{16}$$

the decision rule in the application of the $SG(m)$ test at a $100(1-\alpha)\%$ confidence level is:

$$\begin{aligned} &\textit{If } 0 \leq SG(m) \leq \chi^2_\alpha && \textit{Accept } H_0 \\ &\qquad\qquad \textit{Otherwise} && \textit{Reject } H_0 \end{aligned} \tag{17}$$

4 Proposed Symbolization Procedure

In this section, we propose a standard symbolization map f for the spatial process $\{X_s\}_{s \in S}$. There might be several possible standard symbolization maps, and we invite the reader to do so in order to detect spatial dependence. The procedure we are going to show in this section can be refined in particular cases in which the researcher has a better understanding of the particular process to be studied. The proposed standard symbolization map f is defined as follows: denote by Me the median of the spatial process $\{X_s\}_{s \in S}$ and let

$$Y_s = \begin{cases} 0 & \text{if } X_s \leq Me \\ 1 & \text{otherwise} \end{cases} \tag{18}$$

Now, define the indicator function

$$\mathcal{I}_{s_1 s_2} = \begin{cases} 0 & \text{if } Y_{s_1} \neq Y_{s_2} \\ 1 & \text{otherwise} \end{cases} \tag{19}$$

Then, the standard symbolization map $f : R^m \to \Gamma$ is defined as:

$$f(X_m(s)) = f(X_s, X_{s_1}, \ldots, X_{s_{m-1}}) = (\mathcal{I}_{ss_1}, \mathcal{I}_{ss_2}, \ldots, \mathcal{I}_{ss_{m-1}}) \tag{20}$$

For any localization s set $X_m(s) = (X_s, X_{s_1}, \ldots, X_{s_{m-1}})$. We will denote by $N_s = \{s_1, \ldots s_{m-1}\}$ the $m-1$ nearest neighbors of s. This symbolization procedure consists of comparing at each localization s the value of Y_s with Y_{s_i} for all $s_i \in N_s$. Then if $Y_s = Y_{s_i}$ means that X_s and X_{s_i} are both less than, or greater than, Me and hence σ_s will have the value 1 at the ith-entry.

Notice that, with this symbolization map the set of symbols has cardinality 2^{m-1} and the SG test has $k = n$ (the number of symbols) degrees of freedom. It is straightforward to check that, under the null $H_0 : \{X_s\}_{s\in S}$ i.i.d., all the symbols have the same probability to occur $\frac{1}{2^{m-1}}$, for any continuous process.

5 Finite sample behavior of $SG(m)$ and comparison with other tests for independence

In this section, we examine the finite sample behavior of the $SG(m)$ test. Moreover, we have conducted a power comparison among different non-parametric tests for spatial independence, the τ test of Brett and Pinkse (1997) and the $SBDS$ of de Graaff et al. (2001). In Cliff and Ord (1981), Anselin and Rey (1991), Anselin and Florax (1995), Florax et al (2003) and Florax and de Graaff (2004) different simulations on some of the most popular tests can be found. However, we know of no simulation specifically for the nonparametric tests ($SBDS$ and τ) in a spatial context. In this paper, we resolve an exercise of this type in which we also include the SG test. We will also compare these tests with the classical Moran's I. We first present the tests and then discuss the Monte Carlo simulation results.

Under the alternative, we have considered 4 data generating processes

$$\begin{aligned} &\text{DGP 1} \quad X = (I_n - \rho W)^{-1}\varepsilon, \\ &\text{DGP 2} \quad X = (I_n + \rho W)\varepsilon, \\ &\text{DGP 3} \quad X = (I_n + \rho W)\varepsilon^2, \\ &\text{DGP 4} \quad X = (I_n + \rho W)\varepsilon^3. \end{aligned} \tag{21}$$

where $\varepsilon \sim iid$, $N(0,1)$. DGP1 is an SAR process, DGP2 is an SMA process and DGP3 and DGP4 are two non-linear processes.

Table 1 reports the empirical power of the SG test on different sample sizes. As we can see, the power of our test against dependent processes is certainly satisfactory. Power results rapidly improve as the sample size (lattice dimensions) increases, regardless of the underlying stochastic process. This outcome is highlighted with the intrinsic good properties of the SG test, such as, simplicity, consistency and the absence of restrictive assumptions on the data generating process.

		τ			SG			I			SBDS		
	ρ	0,1	0,5	0,9	0,1	0,5	0,9	0,1	0,5	0,9	0,1	0,5	0,9
DGP1	7 x 7	0,043	0,600	0,996	0,058	0,278	0,919	0,064	0,738	1,000	0,053	0,144	0,871
	12 x 12	0,069	0,938	1,000	0,052	0,583	1,000	0,132	0,984	1,000	0,068	0,245	0,999
	20 x 20	0,091	1,000	1,000	0,070	0,955	1,000	0,219	1,000	1,000	0,045	0,455	1,000
	40 x 40	0,121	1,000	1,000	0,119	1,000	1,000	0,536	1,000	1,000	0,054	0,913	1,000
DGP2	7 x 7	0,059	0,518	0,918	0,051	0,228	0,521	0,086	0,700	0,982	0,066	0,124	0,325
	12 x 12	0,065	0,845	1,000	0,047	0,391	0,912	0,115	0,965	1,000	0,052	0,138	0,534
	20 x 20	0,083	0,968	1,000	0,086	0,796	0,999	0,262	1,000	1,000	0,052	0,182	0,843
	40 x 40	0,109	1,000	1,000	0,113	1,000	1,000	0,496	1,000	1,000	0,055	0,479	0,999
DGP3	7 x 7	0,085	0,716	0,984	0,045	0,280	0,542	0,076	0,728	0,990	0,060	0,158	0,520
	12 x 12	0,094	0,904	1,000	0,057	0,667	0,964	0,123	0,979	1,000	0,052	0,141	0,675
	20 x 20	0,082	0,992	1,000	0,087	0,990	1,000	0,170	1,000	1,000	0,050	0,234	0,893
	40 x 40	0,122	1,000	1,000	0,207	1,000	1,000	0,526	1,000	1,000	0,051	0,278	0,998
DGP4	7 x 7	0,092	0,959	0,999	0,232	0,533	0,636	0,063	0,779	0,986	0,052	0,157	0,532
	12 x 12	0,106	1,000	1,000	0,635	0,967	0,993	0,104	0,968	0,999	0,049	0,142	0,731
	20 x 20	0,142	1,000	1,000	0,985	1,000	1,000	0,193	0,999	1,000	0,036	0,222	0,960
	40 x 40	0,189	1,000	1,000	1,000	1,000	1,000	0,514	1,000	1,000	0,037	0,279	0,999

6 Conclusions

The present paper attempts to analyze limited and noisy data using minimal hypothesis, looking specifically at the assumption of independence. Specifically, we are interested in the competence of a non-parametric approach based on entropy measures, well-established in mainstream Econometrics, but as far as we know al-most unnoticed in a spatial context. Hong and White (2005) present some tests for independence obtained by using entropy measures and provide their asymptotic distribution. The lastest proposal in this line is the G test of Matilla and Ruiz (2008) which offers several advantages with respect to the other nonparametric tests. In line with the suggestions in the above-mentioned work of Matilla and Ruiz, we have proposed a new test for spatial independence, called SG, which relies on the concept of entropy. The last paper also provides the asymptotic standard distribution of an affine transformation of the symbolic entropy under the null of independence. The theoretical distribution allows us to construct a test for independence which is consistent against a broad class of spatial dependences (including those with zero autocorrelation). The empirical size does not differ from the theoretical size, which is an interesting property that guarantees the general applicability and reproducibility of the test. Moreover, the test is invariant under monotonic transformations of data. Invariance makes our procedure very attractive in practice. Most importantly, our test makes no assumptions about the continuous or discrete nature of the data generating process and of its marginal densities, nor it is necessary to specify the weighting matrix. Two final advantages are its computational simplicity and, hence, its short running computational times. We present the main results of a Monte Carlo simulation in which we have included the SG test together with the well-known Moran's I and two other non-parametric spatial independence tests, namely, the $SBDS$ of de Graff et al. (2001) and the τ test of Brett and Pinkse (1997). These results allow to state that the SG has, in general, the right empirical size and good power against several departures from the null hypothesis of independence. Its performance considerably improves with the sample and, in a large sample case, the SG test is fully competitive with respect to the other tests in spite of being less demanding in terms of prior information. Singularly, the SG test is free from the yoke of specifying the weighting matrix.

- Anselin, L. (1988): Spatial Econometrics. Methods and Models. Dordrecht: Kluwer.

- Anselin, L. (2001): Rao's score test in spatial econometrics. Journal of Statistical Planning Inference, 97 113-139.

- Anselin, L., A, Bera, R. Florax and M. Yoon (1996): Simple Diagnostic Tests for Spatial Dependence. Regional Science and Urban Economics, 26, 77-104.

- Anselin, L. and A. Bera, (1998): Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In Ullah, A. and D. Giles (eds.) Handbook of Applied Economic Statistics (pp. 237-289). New York: Marcel Dekker.

- Anselin, L. and R. Florax (1995): Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results. In Anselin, L. and R. Florax (eds.) New Directions in Spatial Econometrics (pp. 21-74). Berlin: Springer.

- Anselin, L and H. Kelejian (1997): Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors. International Regional Science Review, 20, 153-182.

- Anselin, L. and R. Moreno (2003): Properties of tests for spatial error components. Regional Science and Urban Economics, 33 595-618.

- Anselin, L. and S. Rey (1991): Properties of tests for spatial dependence in linear regression models. Geographical Analysis 23 112-131.

- Arbia G, (1989): Spatial data configuration in statistical analysis of regional economics and related problems. Dordrecht: Kluwer Academic Publishers.

- Arbia G, (2006): Spatial Econometrics. Statistical Foundations and Applications to Regional Convergence. Berlin: Springer.

- Baltagi, B., S. Song and W. Koh (2003): Testing panel data regression models with spatial error correlation, Journal of Econometrics, 117 123-150.

- Bavaud, F. (1997): Models for Spatial Weights: A Systematic Approach. Geographical Analysis 30, 1653-171.

- Bera, A and P. Simlai (2004): Testing Spatial Autoregressive Model and a Formulation of Spatial ARCH (SARCH) Model with Applications. Manuscript: Department of Economics, University of Illinois.

- Brett, C. and J. Pinkse (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.

- Brock, W., D. Dechert, J. Scheinkman and B. LeBaron (1996): A Test for Independence Based on the Correlation Dimension. Econometric Review, 15 197-235.

- Brock, W., D. Hsieh and B. LeBaron (1991): Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Evidence. Cambridge: MIT Press.

- Burridge, P. (1980): On the Cliff-Ord test for spatial autocorrelation. Journal of the Royal Statistical Society B 42 107-108.

- Cliff, A. and K. Ord (1981): Spatial Processes. Models and Applications. Pion: London.

- Conley, T. (1999): GMM estimation with cross sectional dependence, Journal of Econometrics, 92 1-45.
- Dacey, M. (1965): A Review of Measures of Contiguity for Two and k-Color Maps. In Berry, B and D. Marble (eds): A Reader in Statistical Geography, pp 479-495. Englewood Cliffs: Prentice-Hall.
- Fingleton, B (2008): A Generalized Method of Moment Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Aberage Errors. Spatial Economic Analysis, 3 27-44.
- Florax, R. and S. Rey (1995): The Impacts of Misspecified Spatial Interaction in Linear Regression Models In Anselin, L. and R. Florax (eds.) New Directions in Spatial Econometrics (pp. 111-135). Berlin: Springer.
- Florax, R., H. Folmer and S. Rey (2003): Specification Searches in Spatial Econometrics: the Relevance of Hendry's Methodology. Regional Science and Urban Economics, 33 557-579.
- Florax, R. and T. de Graaff (2004): The Performance of Diagnostics Tests for Spatial Dependence in Linear Regression Models: A Meta-Analysis of Simulation Studies. In Anselin, L., R. Florax and S. Rey (eds.): Advances in Spatial Econometrics: Methodology, Tools and Applications, pp. 29-65. Berlin: Springer.
- Geary, R. (1954): The Contiguity Ratio and Statistical Mapping. The Incorporated Statistician, 5 115-145
- Getis, A. (2007): Reflections on Spatial Autocorreclation, Regional Science and Urban Economics; 37, 491–496.

- Gould, P. (1970): Is Statistix Inferens the Geographical Name for a Wild Goose? Economic Geography, 46 439-448

- Graaff, de Th., R. Florax and P. Nijkamp (2001): A General Misspecification Tests for Spatial Regression Models: Dependence, Heterogeneity and Nonlinerity. Regional Science and Urban Economics, 41, 225-276.

- Grassberger, P. and P. Procaccia. (1983): Measuring the Strangeness of Strange Attactors. Physica, 9D 189-208.

- Griffith, D. (1996): Some Guideliness for Specifying the Geographic Weight Matrix Contained in Spatial Statistical Models. In Arlinghaus S. (eds): Practical Handbook of Spatial Statistics, pp 65-82. Boca Raton: CRC.

- Haining, R. (2003): Spatial Data Analysis: Theory and Practice. Cambridge: Cambridge University Press

- Hong, Y. and H. White, H. (2005): Asymptotic distribution theory for nonparametric entropy measures of serial dependence. Econometrica 73, 837-901.

- Joe, H. (1989a): Relative Entropy Measures of Multivariate Dependence. Journal of the American Statistical Association, 84, 157-164.

- Joe, H.(1989b): Estimation of Entropy and Other Functionals of a Multivariate Density. Annals of the Institute of Statistical Mathematics. 41, 683-697.

- Kelejian, H. and I. Prucha (1999): A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. International Economic Review, 40 509-533.

- Kelejian, H. and I. Prucha (2001): On the asymptotic distribution of the Moran I test statistic with applications, Journal of Econometrics, 104 219-257.

- Kelejian H. and I. Prucha (2004): Estimation of systems of spatially interrelated cross sectional equations, Journal of Econometrics 118 27-50.

- Kelejian H. and I. Prucha (2007): HAC estimation in a spatial framework Journal of Econometrics, 140 131-154

- Kelejian H. and D. Robinson (1992): Spatial autocorrelation: A new computationally simple test with an application to per capita country police expenditures. Regional Science and Urban Economics, 22 317-333.

- Kelejian H. and P. Robinson (1993): A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science, 72 297-312.

- Kelejian, H and D. Robinson (1995): Spatial correlation: a suggested alternative to the autoregressive model. In L. Anselin and R. Florax (eds): New Directions in Spatial Econometrics (pp. 75-95), Springer-Verlag, Berlin.

- King, M. (1981): A Small Sample Property of Cliff-Ord Test for Spatial Correlation. Journal of the Royal Statistical Sciety B, 43 263-264.

- King, M. (1985): A Point Optimal Test for Autoregressive Disturbances. Journal of Econometrics, 27 21-37.

- Knox, E. (1964): The Detection of Space-Time Interactions. Applied Statistics, 13 25-29.

- Lehmann, E.L. (1986): *Testing Statistical hypothesis*. John Wiley & Sons, New York.

- Leung Y., C. Mei and W. Zhang (2003): Statistical Tests for Local Patterns of Spatial Association. Environment and Planning A. 35: 725-744.

- Matilla-García M., and Ruiz Marín, M. (2008): A non-parametric independence test using permutation entropy. Journal of Econometrics, 144 139–155.

- Moran, P. (1950): Notes on Continuous Stochastic Phenomena. Biometrika, 37 17-23.

- Nijkamp, P. and A. Reggiani (1998): The Economics of Complex Spatial Systems. JAI Pres: London.

- Paelinck, J. and L. Klaassen (1979): Spatial Econometrics. Farnborough: Saxon House.

- Pinkse, J. (1998): A Consistent Nonparametric Test for Serial Dependence. Journal of Econometrics, 84 205-231.

- Pinkse, J. (2004): Moran-flavoured tests with nuisance parameters: examples. In: Anselin, L., R. Florax, and S. Rey (eds.), New Advances in Spatial Econometrics (pp. 67-77). Springer, New York.

- Pinkse, J., M. Slade and C. Brett (2002): Spatial Price Competition: A Semiparametric Approach. Econometrica, 70 1111-1153.

- Rohatgi, V.K. (1976): An Introduction to Probability Theory and Mathematical Statistics. John Wiley & Sons, New York.

- Saavedra, L. (2003): Tests for spatial lag dependence based on method of moments estimation. Regional Science and Urban Economics, 33 27-58.

- Sen, A. (1976): Large Sample-Size Distribution of Statistics Used in Testing for Spatial Correlation. Geographical Analysis, 8 175-184.

- Tiefelsdorf, M. (2000): Modelling Spatial Processes. The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I. Berlin: Springer.

- Tiefelsdorf, M and B. Boots (1995): The Exact Distribution of Moran's I. Environment and Planning A, 27 985-999.

- Tobler, W. (1970): A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, 46 234-240.