



Munich Personal RePEc Archive

A Non-Parametric Approach to Spatial Causality

Herrera Gómez, Marcos and Ruiz Marín, Manuel and Mur Lacambra, Jesús and Paelinck, Jean

2010

Online at <https://mpra.ub.uni-muenchen.de/36768/>

MPRA Paper No. 36768, posted 19 Feb 2012 05:50 UTC

A Non-Parametric Approach to Spatial Causality

Marcos Herrera(*) Manuel Ruiz(**) Jesús Mur(*) Jean Paelinck(***)

() Department of Economic Analysis. University of Zaragoza.*

E-mail: jmur@unizar.es; mherreragomez@gmail.com

*(**) Department of Quantitative Methods and Computing.*

Politechnical University of Cartagena.

E-mail: manuel.ruiz@upct.es

*(***) School of Public Policy. George Mason University.*

E-mail: j.paelinck@planet.nl

Abstract

The purpose of this paper is to show the capacity of a new non-parametric test based on symbolic entropy and symbolic dynamics to deal with the detection of linear and non-linear spatial causality. The good performance of the new test in detecting spatial causality and causal weighting matrix is notable and gives rise to an expectation that it may form a adequate tool for constructive specification searches.

Keywords: Causality; Spatial Dependence; Spatial Weight Matrices.

JEL Classification: C21; C50; R15

1 Introduction

Detection of cause-effects relationships among variable has been one of the fundamental questions of most part of natural or social sciences, including Economics. The bibliometric study of Hoover (2004, p.4) is very illustrative: 70% of the articles in the *JSTOR* archives published in 2001, contain words ‘*in a causal family (“cause”, “causes”, “causal”, “causally” or “causality”)*’. The percentage increases up to the 80% if the search is restricted only to econometric articles. It is clear that causality is one of the leading topics in mainstream Economics and Econometrics. In contrast, the Index of Lesage and Pace (2009) textbook on Spatial Econometrics contains almost 1,000 headwords, none of which is related to the Hoover’s causal family. Exactly the same can be said with respect to other textbooks published in the field such as Paelinck and Klaassen (1979), Anselin (1988), Upton and Fingleton (1985), Anselin and Florax (1995), Tiefelsdorf (2000), Griffith (2003), Anselin, Florax and Rey (2004), Getis, Mur and Zoller (2004) or Arbia (2006). This silence is striking and hardly justifiable.

The traditional approach to causality (i.e., Suppes, 1970) insists in the idea of temporal precedence: the cause must occur before the effect. Granger (1980) adds a second fundamental clause: the variable supposed to be the cause must contain information about the effect that is unique, and is in no other variable. The consequence is that the causal variable should help to forecast the effect variable, leading to the concept of ‘*incremental predictability*’ as a quantifiable, and successful, measure of causality. This is the same idea made by Wiener (1956): ‘*For two simultaneously measured signals, if we can predict the first signal better by using the past information from the second one than by using the information without it, then we call the second signal causal to the first one*’.

In a typical spatial econometric problem we have a single cross-section collection of contemporaneous data, without time perspective. Forecasting is not a hot topic here where, put it very rude, the main problem is explaining the spatial distribution of some variable according to different elements. The accent is in ‘explaining’ not in ‘forecasting’ which appears to exclude (unreasonably from our point of view) causality from the toolbox.

A similar situation occurs in other disciplines (such as physics, biology, climatology, to mention a few) where there is a sharp interest not only in detecting synchronized states (that is, coupled systems highly correlated between their internal dynamical states; i.e., Manrubia *et al*, 2004) but also in identifying drive-response relationships. The last point amounts to identify the ‘arrow’ of causality in the evolution of the interacting system. The generalization of the Granger-Wiener approach is one strand of the solution, not very reliable when the relations are nonlinear (Ancona *et al*, 2004). Measures based on the Theory of Information appear to be more robust. Schreiber (2000) proposes a non-parametric method for measuring causal information

transfer between systems, called *transfer entropy*, which is simple and powerful (see, also Marschinski and Kantz, 2002, or Dicks and Panchenko, 2006, for similar works). Our proposal, also nonparametric, is close to this line of reasoning and appears to be well adapted to a typical spatial econometrics application.

The method that we present is based on permutation entropy (see Joe, 1989a and b, Hong and White, 2005, and references therein), a flexible non-parametric technique aimed at finding regular patterns in large collections of data making few assumptions. Matilla and Ruiz (2008) introduce symbolic dynamics in this framework, with the purpose of summarizing the fundamental information that exist in a time series, and the symbolic entropy as a way of quantifying this volume of information. The authors also obtain a well-behaved test of non-serial dependence that Lopez *et al.* (2010) extend to the spatial case. In continuation, we adapt these techniques to the problem of how to identify causal relationships in a cross-sectional spatial context.

Section 2 introduces the notation, definition and basic elements of our approach. In Section 3 we present the test of spatial causality which is based on the comparison of two measures of conditional entropy. In order to do this we need to use a well-defined symbolization procedure. Finally, the test, that has a certain flavor of the original Granger-Wiener framework, is solved using a bootstrap procedure. Section 4 focuses on the subtle question of selecting the most adequate spatial structure for solving the test. This part of the discussion is similar to the problem of defining the relevant lag span in a time series context. Section 5 presents the results of our test in a large Monte Carlo, using linear and nonlinear relations between the variables. Main conclusions appear in the sixth section.

2 Preliminaries

Let $m \in \mathbb{N}$ 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:

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

where s_1, s_2, \dots, s_{m-1} are the $m - 1$ nearest neighbors to s_0 , which are ordered from lesser to higher Euclidean distance with respect to location s_0 . If two or more locations are equidistant to s_0 we choose them in an anticlockwise manner. In formal terms, s_1, s_2, \dots, s_{m-1} are the $m - 1$ nearest neighbors to s_0 satisfying the following two conditions:

$$\begin{aligned} (a) \quad & \rho_1^0 \leq \rho_2^0 \leq \dots \leq \rho_{m-1}^0, \\ (b) \quad & \text{and if } \rho_i^0 = \rho_{i+1}^0 \text{ then } \theta_i^0 < \theta_{i+1}^0 \end{aligned} \tag{1}$$

Notice that condition (b) is a technical condition that ensures the uniqueness of $X_m(s_0)$ for all $s \in S$ in the case in which two neighbors are at the same distance of s_0 . We will call $X_m(s)$ an m – surrounding of point s .

Let $\Gamma_n = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$ be a set of n symbols. Now assume that there is a map

$$f : \mathbb{R}^m \rightarrow \Gamma$$

defined by $f(X_m(s)) = \sigma_{j_s}$ with $j_s \in \{1, 2, \dots, n\}$. 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*. We will say that the symbol $\sigma \in \Gamma$ is *admissible* for the spatial process $\{X_s\}_{s \in S}$ if and only if $f(X_m(s)) = \sigma$ for some $s \in S$.

Denote by

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

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:

$$p(\sigma) := p_\sigma = \frac{\# \{s \in S | s \text{ is of } \sigma \text{ - type}\}}{|S|} \quad (2)$$

where by $|S|$ we denote the cardinality of 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 Shannon's entropy of the n distinct symbols as follows:

$$h_m(X) = - \sum_{\sigma \in \Gamma} p_\sigma \ln(p_\sigma). \quad (3)$$

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

Consider now a k - *dimensional* spatial process $\{Z_s = (X_{1s}, X_{2s}, \dots, X_{ks})\}_{s \in S}$ and a fix embedding dimension m . Let $\Gamma_k = \Gamma \times \Gamma \cdots \times \Gamma$ the direct product of k copies of Γ . Let $\eta_{i_1, i_2, \dots, i_k} = (\sigma_{i_1}, \sigma_{i_2}, \dots, \sigma_{i_k}) \in \Gamma_k$. Then we will say that s is of $\eta_{i_1, i_2, \dots, i_k}$ - *type* for Z if and only if s is of σ_{i_j} - *type* for X_{j_s} for all $j = 1, 2, \dots, k$. Then under this context we can define symbolic entropy for the k - *dimensional* spatial process $\{Z_s\}_{s \in S}$ as:

$$h_m(Z) = - \sum_{\eta \in \Gamma_k} p_\eta \ln(p_\eta) \quad (4)$$

Denote the conditional entropy of Y conditioned to the occurrence of symbol σ^x in X by:

$$h_m(Y|\sigma^x) = - \sum_{\sigma^y \in \Gamma} p(\sigma^y|\sigma^x) \ln(p(\sigma^y|\sigma^x)). \quad (5)$$

Then we define conditional symbolic entropy of Y_s given X_s :

$$h_m(Y|X) = - \sum_{\sigma^x \in \Gamma} p(\sigma^x) h_m(Y_s|\sigma^x) \quad (6)$$

as the average of symbolic entropies with respect to conditional pmf's.

3 The Test

Let $\{X_s\}_{s \in S}$ and $\{Y_s\}_{s \in S}$ be two real valued spatial processes. Let

$$\mathcal{W}(X, Y) = \{W_i | i \in I\} \quad (7)$$

be a set of weighting matrices determining all possible spatial causal relations between the two spatial processes, where I is a set of indexes. We will call the set $\mathcal{W}(X, Y)$ *causal spatial structure* from X to Y .

Denote by

$$\mathcal{X}_{\mathcal{W}} = \{W_i X | W_i \in \mathcal{W}(X, Y)\} \quad (8)$$

the set of spatial lags of X given by all the causal spatial structures from X to Y . Our definition of causality is based on information theoretic arguments.

Definition: We will say that $\{X_s\}_{s \in S}$ does not cause $\{Y_s\}_{s \in S}$ under the causal spatial structure $\mathcal{W}(X, Y)$ if

$$h_m(Y) = h_m\{Y | \mathcal{X}_{\mathcal{W}}\} \quad (9)$$

Therefore we propose to perform a non-parametric one-sided test for the following null hypothesis

$$H_0 : \{X_s\}_{s \in S} \text{ does not cause } \{Y_s\}_{s \in S} \text{ under the causal spatial structure } \mathcal{W}(X, Y)$$

with the following statistic:

$$\hat{\delta}(\mathcal{W}) = \hat{h}_m(Y) - \hat{h}_m\{Y | \mathcal{X}_{\mathcal{W}}\} \quad (10)$$

If $\mathcal{X}_{\mathcal{W}}$ does not contain extra information about Y then $\hat{\delta}(\mathcal{W}) = 0$, otherwise $\hat{\delta}(\mathcal{W}) > 0$. In order for the bootstrapped test to be asymptotically independent of

the bootstrap *DGP* we have to ensure that the bootstrap *DGP* respects the null hypothesis of no causality. To this end we have resampled $\{X_s\}_{s \in S}$ and $\{Y_s\}_{s \in S}$ independently rather than jointly, since the pairwise resampling may preserve the underlying causality in the bootstrapped data. Note that the dependence structure present in the original data is unfortunately lost.

The bootstrap test procedure, with a number B of bootstrap replications, is composed of the following steps:

1. Compute the value of the statistic $\hat{\delta}(\mathcal{W})$ for the original samples $\{X_s\}_{s \in S}$ and $\{Y_s\}_{s \in S}$.
2. By resampling $\{X_s\}_{s \in S}$ and $\{Y_s\}_{s \in S}$, obtain two bootstrapped series $\{X_s(b)\}_{s \in S}$ and $\{Y_s(b)\}_{s \in S}$, where b indicates the number of bootstrapped sample.
3. For the bootstrapped samples estimate the bootstrap realization of the statistic of interest denoted by:

$$\hat{\delta}^{(b)}(\mathcal{W}) = \hat{h}_m(Y(b)) - \hat{h}_m(Y(b) | \mathcal{X}_{\mathcal{W}}(b)) \quad (11)$$

4. Repeat $B - 1$ times steps 2 and 3 to obtain B bootstrap realizations of the statistic, $\{\hat{\delta}^{(b)}(\mathcal{W})\}_{b=1}^B$.
5. Compute the bootstrap *p*-value:

$$p_{\text{boots} - \text{value}}(\hat{\delta}(\mathcal{W})) = \frac{1}{B} \sum_{b=1}^B 1(\hat{\delta}^{(b)}(\mathcal{W}) > \hat{\delta}(\mathcal{W})) \quad (12)$$

where $1(\cdot)$ is the indicator function which assigns 1 to a true statement and 0 otherwise.

6. Reject the null hypothesis of $\{X_s\}_{s \in S}$ does not cause and $\{Y_s\}_{s \in S}$ under the spatial structure $\mathcal{W}(X, Y)$ if

$$p_{\text{boots} - \text{value}}(\hat{\delta}(\mathcal{W})) < \alpha \quad (13)$$

for a nominal size α .

3.1 A Proposal of Symbolization

Now we propose a particular symbolization map f for the spatial process $\{X_s\}_{s \in S}$. There might be several possible symbolization maps. Therefore, this novel framework is adaptable to the necessities of the problem at hand, and so the procedure below

can be refined in accordance with particular cases for which the researcher has a better understanding of the process to be studied. The proposed symbolization map f is defined as follows: denote by Me the median of the spatial process $\{X_s\}_{s \in S}$ and let

$$\gamma_s = \begin{cases} 0 & \text{if } X_s \leq Me \\ 1 & \text{otherwise} \end{cases} \quad (14)$$

Now, we define the indicator function

$$\mathcal{I}_{s_1 s_2} = \begin{cases} 0 & \text{if } \gamma_{s_1} \neq \gamma_{s_2} \\ 1 & \text{otherwise} \end{cases} \quad (15)$$

For any localization s , set $X_m(s) = (X_s, X_{s_1}, \dots, X_{s_{m-1}})$. We denote by $N_s = \{s_1, \dots, s_{m-1}\}$ the $m-1$ nearest neighbors of s . This symbolization procedure consists of comparing at each localization s the value of γ_s with γ_{s_i} for all $s_i \in N_s$. Thus, that $\gamma_s = \gamma_{s_i}$ means that X_s and X_{s_i} are both less than, or greater than, Me . Therefore the value $\varphi(s) = \sum_{s_i \in N_s} \mathcal{I}_{ss_i}$ gives us at each location $s \in S$ the number of neighbors of s that agree with X_s to be either or below the median Me .

Then, the symbolization map $f : \mathbb{R}^m \rightarrow \Gamma$ is defined as:

$$f(X_m(s)) = f(X_s, X_{s_1}, \dots, X_{s_{m-1}}) = \varphi(s) = \sum_{s_i \in N_s} \mathcal{I}_{ss_i} \quad (16)$$

where $\Gamma = \{0, 1, 2, \dots, m-1\}$.

4 Detection of Causal Weighting Matrices

Let $\{X_s\}_{s \in S}$ and $\{Y_s\}_{s \in S}$ be two spatial processes such that X causes Y . Let $\mathcal{W}(X, Y)$ be the causal spatial structure from X to Y . This section is devoted to detect which $W \in \mathcal{W}(X, Y)$ is the most significant revealing the causal spatial structure.

Let \mathcal{K} be any subset of Γ and let $W \in \mathcal{W}(X, Y)$. Then we define

$$\mathcal{K}_W^X = \{\sigma^x \in \mathcal{K} \mid \sigma^x \text{ is admissible for } WX\}. \quad (17)$$

We will denote by Γ^X the set of symbols that are admissible for $\{X_s\}_{s \in S}$. Let $W_0 \in \mathcal{W}(X, Y)$ be the most significant weighting matrix revealing the causal spatial structure from X to Y . Given the spatial process $\{Y_s\}_{s \in S}$ there exists a subset $\mathcal{K} \subseteq \Gamma$ such that $p(\mathcal{K}_{W_0}^X \mid \sigma^y) > p(\mathcal{K}'^X \mid \sigma^y)$ for all $\mathcal{K}' \subseteq \Gamma, W \in \mathcal{W}(X, Y)$ and $\sigma^y \in \Gamma^Y$. Therefore

$$\begin{aligned}
h_m(W_0X|Y) &= - \sum_{\sigma^y \in \Gamma^Y} p(\sigma^y) \left[\sum_{\sigma^x \in \mathcal{K}_{W_0}^X} p(\sigma^x|\sigma^y) \ln(p(\sigma^x|\sigma^y)) \right] \leq & (18) \\
&\leq - \sum_{\sigma^y \in \Gamma^Y} p_{\sigma^y} \left[\sum_{\sigma^x \in \mathcal{K}'_W^X} p(\sigma^x|\sigma^y) \ln(p(\sigma^x|\sigma^y)) \right] = h_m(WX|Y)
\end{aligned}$$

Thus we have proved the following theorem.

Theorem 4.1: *Let $\{X_s\}_{s \in S}$ and $\{Y_s\}_{s \in S}$ be two spatial process. Assume that X_s causes Y_s under the causal spatial structure $\mathcal{W}(X, Y)$. For a fixed embedding dimension $m > 2$, with $m \in \mathbb{N}$, if the most important weighting matrix revealing the causal spatial structure from X to Y is $W_0 \in \mathcal{W}(X, Y)$ then*

$$h_m(W_0X|Y) = \min_{W \in \mathcal{W}(X, Y)} \{h_m(WX|Y)\}. \quad (19)$$

5 Monte Carlo Simulations

In this section we present information about the finite sample behavior of the two complementary techniques, developed in the previous sections to deal with the question of causality in a spatial context.

Section 5.1 focuses on the application of Theorem 4.1; that is, on the measure of the conditional entropy of (19) as a criterion for selecting the most important weighting matrix in a causal relation. The concern of Section 5.2 is with the performance of the causality test of (10). As said, both problems are connected: in first place a weighting matrix must be chosen then, and conditional on this selection, a causality test can be solved. This is the order of the discussion that follows.

5.1 Selecting a Weighting Matrix

As said, in this section we present some evidence on the performance of Theorem 4.1 when applied to linear and nonlinear processes under different scenarios.

The data-generating processes (*DGPs*) studied are the following

$$\begin{aligned}
DGP1 : Y &= \rho W X + \nu & X &= \varepsilon \\
DGP2 : Y &= 1/(\rho W X + \nu) & X &= \varepsilon \\
DGP3 : Y &= (\rho W X + \nu)^5 & X &= \varepsilon \\
DGP4 : Y &= \sin(\rho W X + \nu) & X &= \varepsilon
\end{aligned}$$

where ε and ν are normal standard distributed.

To evaluate the performance of the nonparametric method in finite samples, we compute 1000 Monte Carlo replications of each model, and we consider 5 different

matrices W_1, W_2, W_3, W_4 and W_5 , each of them generated over an irregular (random) lattice. We estimate $h(WX|Y)$ for sample size $R = 400$, three values of $\rho = 0.5, 1$ and 2 and three values of m , namely, $m = 4$ (thus only 16 symbols are used to obtain a conclusion about the spatial structure of the spatial process Y), $m = 5$ (25 symbols are used) and $m = 6$ (36 symbols are used).

In the following table we present the results of the average value of the $h(W_iX|Y)$ for each $i = 1, 2, 3, 4$ and 5 for a given model, with average computed over the total number of Monte Carlo replications.

Table 1: Simulations of Conditional Entropy

<i>DGP 1</i>	$\rho = 0.5$			$\rho = 1$			$\rho = 2$		
<i>m</i>	4	5	6	4	5	6	4	5	6
W_1	1,2983	1,4746	1,6109	1,2798	1,4580	1,5955	1,1911	1,3723	1,5085
W_2	1,2945	1,4767	1,6174	1,2946	1,4776	1,6163	1,2949	1,4772	1,6180
W_3	1,2987	1,4771	1,6132	1,2966	1,4778	1,6120	1,2987	1,4774	1,6151
W_4	1,2991	1,4735	1,6145	1,2998	1,4736	1,6141	1,3000	1,4742	1,6141
W_5	1,2920	1,4790	1,6176	1,2904	1,4775	1,6175	1,2920	1,4773	1,6178

<i>DGP 2</i>	$\rho = 0.5$			$\rho = 1$			$\rho = 2$		
<i>m</i>	4	5	6	4	5	6	4	5	6
W_1	1,2991	1,4752	1,6116	1,2863	1,4640	1,6017	1,2254	1,4091	1,5451
W_2	1,2942	1,4765	1,6178	1,2945	1,4777	1,6166	1,2950	1,4774	1,6178
W_3	1,2986	1,4776	1,6133	1,2968	1,4780	1,6121	1,2989	1,4775	1,6149
W_4	1,2988	1,4736	1,6144	1,3001	1,4733	1,6140	1,2997	1,4736	1,6147
W_5	1,2919	1,4789	1,6172	1,2904	1,4773	1,6181	1,2918	1,4776	1,6183

<i>DGP 3</i>	$\rho = 0.5$			$\rho = 1$			$\rho = 2$		
<i>m</i>	4	5	6	4	5	6	4	5	6
W_1	1,2968	1,4757	1,6142	1,2779	1,4595	1,5957	1,1852	1,3772	1,5162
W_2	1,2913	1,4863	1,6211	1,2904	1,4870	1,6202	1,2904	1,4863	1,6204
W_3	1,2984	1,4792	1,6189	1,2993	1,4814	1,6190	1,2998	1,4784	1,6198
W_4	1,2977	1,4795	1,6157	1,2981	1,4798	1,6154	1,2975	1,4785	1,6140
W_5	1,2950	1,4761	1,6189	1,2961	1,4735	1,6197	1,2957	1,4763	1,6195

<i>DGP 4</i>	$\rho = 0.5$			$\rho = 1$			$\rho = 2$		
<i>m</i>	4	5	6	4	5	6	4	5	6
W_1	1,2985	1,4747	1,6111	1,2833	1,4600	1,5972	1,2442	1,4115	1,5403
W_2	1,2943	1,4769	1,6178	1,2945	1,4776	1,6162	1,2949	1,4769	1,6181
W_3	1,2986	1,4772	1,6133	1,2966	1,4778	1,6121	1,2990	1,4776	1,6148
W_4	1,2992	1,4735	1,6143	1,3000	1,4736	1,6139	1,3000	1,4744	1,6147
W_5	1,2919	1,4791	1,6175	1,2903	1,4773	1,6176	1,2916	1,4778	1,6178

As it is shown in Table 1, conditional entropy clearly detects of causal weighting matrix with $m = 6$. According to the simulation summarized, it can be observed that, for $\rho = 1$ and 2 , regardless m , the minimum entropy is reached at the expected

(correct) causal weighting matrix. Also, the detection of causal weighting matrix is more apparent as m increase. The last observation is also expected because, as m grows, conditional entropy is evaluated on an increasing number of symbols and so a finer search is carried out.

To conclude this section we are going to simulate the following mixed data generating process:

$$DGP5 : Y = 2\rho W_1 X + \rho W_2 X + \nu \quad X = \varepsilon$$

Table 2: Simulations of Conditional Entropy (DGP5)

	$\rho = 0.5$			$\rho = 1$			$\rho = 2$		
m	4	5	6	4	5	6	4	5	6
W_1	1.2985	1,4747	1,6111	1,2833	1,4600	1,5972	1,2442	1,4115	1,5403
W_2	1.2943	1,4769	1,6178	1,2945	1,4776	1,6162	1,2949	1,4769	1,6181
W_3	1.2986	1,4772	1,6133	1,2966	1,4778	1,6121	1,2990	1,4776	1,6148
W_4	1.2992	1,4735	1,6143	1,3000	1,4736	1,6139	1,3000	1,4744	1,6147
W_5	1.2919	1,4791	1,6175	1,2903	1,4773	1,6176	1,2916	1,4778	1,6178

These results are very interesting because they stress the efficiency of the conditional entropy indicator to select the most relevant spatial structure. Data of variable Y have been obtained using two different weighting matrices, W_1 and W_2 . The most important matrix is W_1 because, in $DGP5$, the coefficient associated to this matrix is higher (doubles) than that associated to W_2 . As can be seen in the table above, the conditional entropy indicator, on average, always selects the W_1 matrix (which is, indeed, the most influential) for high values of ρ and/or high values of m , the embedding dimension. Similar results are obtained with other combinations of weights.

5.2 The Causality Test

In continuation we present Monte Carlo results in relation to the performance of the $\hat{\delta}(\mathcal{W})$ statistic of (10) when applied to the problem of testing for causality in linear and nonlinear spatial processes. In order to conduct size and power experiments we have maintained the same collection of models of the previous section:

$$\begin{aligned}
 DGP1 : Y &= \nu & X &= \varepsilon \\
 DGP2 : Y &= (I - 0.5W)^{-1}v & X &= (I - 0.5W)^{-1}\varepsilon \\
 DGP3 : Y &= \rho W X + \nu & X &= \varepsilon \\
 DGP4 : Y &= 1/(\rho W X + \nu) & X &= \varepsilon \\
 DGP5 : Y &= (\rho W X + \nu)^5 & X &= \varepsilon \\
 DGP6 : Y &= \sin(\rho W X + \nu) & X &= \varepsilon
 \end{aligned}$$

where ε and v are normal standard distributed and independent among them.

$DGP's$ 1 – 2 will be used to study the size of the test while $DGP's$ 3 – 6 will be used to study the power performance under linear and nonlinear processes.

Table 3 shows the empirical size of the statistics for the small sample size at usual nominal levels. In general, empirical size results are acceptable. The $\hat{\delta}(\mathcal{W})$ test show

Table 3: Size performance of the $\hat{\delta}(\mathcal{W})$ statistic at 5% significance level

	$R = 100$	$R = 400$	$R = 1000$
m	4	5	6
<i>DGP1</i>	5.5	5.5	5.2
<i>DGP2</i>	6.3	8.2	7.3

a stable behavior around nominal levels for model 1, and values slightly higher but acceptable for model 2.

Table 4: Power performance of the $\hat{\delta}(\mathcal{W})$ statistic in percentage

	$\rho = 0.5$			$\rho = 1$			$\rho = 2$		
m	4	5	6	4	5	6	4	5	6
R	100	400	1000	100	400	1000	100	400	1000
<i>DGP3</i>	6	10	13.5	20	39	85	73	99	100
<i>DGP4</i>	6	9.5	12	10	26.5	51.5	44	91	100
<i>DGP5</i>	6	10	12	22.5	39	69	74.5	99	100
<i>DGP6</i>	8	10.5	14	16	35	66	38.5	95.5	100

Table 4 reports the empirical power of the $\hat{\delta}(\mathcal{W})$ test on different sample sizes. As we can see, when $\rho = 2$, the power of our test against dependent processes is certainly satisfactory. For $\rho = 1$, the power of the test rapidly improves as the sample size and m increases.

6 Conclusions

The purpose of this paper was twofold. In first place, we would like to claim for the importance of the question of causality also in a spatial context. This is one of the leading topics in mainstream Econometrics, surprisingly absent in the Spatial Econometrics agenda.

We contribute to this update with a nonparametric statistic that, specifically, tests for the existence of causality in a pair of variables. This test, which is not restricted to a spatial context, assesses the likelihood ‘arrow’ of causality between the two variables using a measure of conditional entropy and a bootstrapping. Furthermore, we complete the discussion with the development of a technique aimed at selecting the most influential spatial weighting matrix in a causal relation. According to our knowledge, there are few guidelines in relation to how choosing the spatial lag in a given model.

In the paper we present some Monte Carlo evidence of the performance of our proposals in a context of finite samples. Overall, it must be acknowledged that spatial causality it is a difficult question though our preliminary results seem encouraging and claim for further developments.

References

- [1] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht, Kluwer.
- [2] Anselin, L. and R. Florax (eds.) (1995). *New Directions in Spatial Econometrics*. Berlin, Springer.
- [3] Anselin, L., Florax, R. and S. Rey (eds.) (2004). *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin, Springer.
- [4] Ancona, N., D. Marinazzo and S. Stramaglia (2004): Radial Basis Function Approach to Nonlinear Granger Causality in Time Series, *Physical Review E* **70** 056221.
- [5] Arbia, G. (2006). *Spatial Econometrics. Statistical Foundations and Applications to Regional Convergence*. Berlin, Springer.
- [6] Diks, C. and V. Panchenko (2006): A New Statistic and Practical Guidelines for Nonparametric Granger Causality Testing. *Journal of Economic Dynamics and Control* **30**, 1647-1669.
- [7] Getis, A., Mur, J. and H. Zoller (eds.) (2004). *Spatial Econometrics and Spatial Statistics*. Houndmills, Macmillan.
- [8] Griffith, D. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin, Springer.
- [9] Hong, Y. and H. White, H. (2005): Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence. *Econometrica* **73**, 837-901.
- [10] Joe, H. (1989a): Relative Entropy Measures of Multivariate Dependence. *Journal of the American Statistical Association* **84**, 157-164.
- [11] Joe, H. (1989b): Estimation of Entropy and Other Functionals of a Multivariate Density. *Annals of the Institute of Statistical Mathematics* **41**, 683-697.
- [12] Lesage, J. and Pace, K. (2009): *Introduction to Spatial Econometrics*. London, Chapman & Hall/CRC.
- [13] López, F., Matilla-García, M., Mur, J., and M. Ruiz Marín (2010): A Non-Parametric Spatial Independence Test Using Symbolic Entropy. *Regional Science and Urban Economics* **40**, 106-115.

- [14] Manrubia, S., A. Mikhailov and D. Zannette (2004): *Emergence of Dynamical Order: Synchronization Phenomena in Complex Systems*. Singapore: World Scientific Publishing.
- [15] Marschinski, R. and H. Kantz (2002): Analysing the Information Flow between Financial Time Series- An Improved Estimator of Transfer Entropy. *European Physical Journal B* **30**, 275-281.
- [16] Matilla-García, M. and M. Ruiz Marín (2008): A Non-Parametric Independence Test Using Permutation Entropy. *Journal of Econometrics* **144**, 139-155.
- [17] Paelinck, J. and L. Klaassen (1979): *Spatial Econometrics*. Farnborough: Saxon House.
- [18] Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- [19] Schreiber, T. (2000): Measuring Information Transfer. *Physical Review Letters* **85**, 461-464.
- [20] Suppes, P. (1970): *A probabilistic Theory of Causality* (Acta Philosophica Fennica. XXIV). Amsterdam: North-Holland.
- [21] Tiefelsdorf, M. (2000): *Modelling Spatial Processes. The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I*. Springer: Berlin.
- [22] Upton, G. and B. Fingleton (1985): *Spatial Data Analysis by Example*, Vol. 1. New York, Wiley.
- [23] Wiener, N. (1956): The Theory of Prediction. In Beckenbach, E. (ed.): *Modern Mathematics for Engineers*. New York: McGraw.