Some examples of the use of a new test for spatial causality

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

Manuel Ruiz(**)
Department of Quantitative Methods and Computing Technical University of Cartagena

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

Abstract

The purpose of this paper is to illustrate the use of a new non-parametric test for spatial causality, between two series, in a pure cross-section. The basis of the proposal, as in the case of the Granger test, measures the quantity of information added by the variable supposed to be cause with respect to the second variable, supposed to be the effect. The test is robust to possible nonlinearities that may affect the relation. The performance of the new test is adequate under a large variety of situations. Moreover, the test is simple to obtain because we only need some basic probabilities. In order to facilitate the interpretation of the test, we include three examples. Two of them use simulated data whereas the third corresponds to the debate between unemployment and migration.

Keywords: Causality; Symbolic Entropy; Unemployment versus Migration.

JEL Classification: C21, C50, R15

AMS Classification: 62F03

(*) mherreragomez@gmail.com
(**) manuel.ruiz@upct.es
(***) jmur@unizar.es

Algunos ejemplos sobre el uso de un nuevo test de causalidad espacial

Resumen

El objetivo del trabajo es presentar un test no paramétrico dirigido a contrastar no causalidad entre dos series espaciales utilizando un único corte transversal. El contraste se basa, al igual que el test de Granger, en una medida de la cantidad de información aportada por la variable supuesta causa con respecto a la variable supuesta efecto. El test es robusto a posibles no linealidades que puedan afectar la relación y su comportamiento es el adecuado para una gran variedad de casos. Además, su obtención es simple porque únicamente necesitamos estimar una serie básica de frecuencias. Incluimos tres ejemplos para facilitan su interpretación. En dos de ellos utilizamos datos simulados mientras que el tercero trata sobre la relación entre desempleo y migración.

Palabras clave: Causalidad; Entropía simbólica; Desempleo frente a Migración.

Clasificación JEL: C21, C50, R15

Clasificación AMS: 62F03

1. Introduction

Detection of causal relationships among variables has been one of the key challenges of econometrics since the seminal works of Haavelmo (1943) and Jan Tinbergen (1951). Two main tendencies can be appreciated after them: the stochastic approach, emphasizing the principles of regularity and asymmetry between cause and effect, and the structural approach where identification is the basic premise. The contribution of the Cowles Commission was fundamental in the development of the last one whereas Granger causality test and the vector autoregression approach are ultimate results of the first. The process is well documented by Morgan (1991). Hoover (2004) puts numbers into the discussion with a very impressive result: 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 80% if the search is restricted only to articles in econometrics.

Indeed, thereis a huge literature devoted to the topic (see, for example, Bauwens et al, 2006, and references from there), but causality is not as simple as it seems (Pearl, 2009). The list of problems is recurrent: symmetry versus asymmetry, invariance and identificability, common causes, counterfactuals, observational studies, etc. The principle of temporal precedence constitutes a helpful assumption, but is also a restriction that limits the notion of causality. In this sense, the absence of an explicit time ordering of spatial data partly explains the null impact of this topic on the most popular textbooks of Spatial Econometrics. Our impression is that the spatial econometric methodology cannot avoid the notion of causality even if the data are purely cross-sectional (see Weinhold and Nair, 2001, Hurlin and Venet, 2001, Hood et al, 2008, or Tervo, 2009, for the case of spatial panel data).

Our proposal is in line with the notion of '*incremental predictability*' as stated by Granger (1980): the variable supposed to be the cause must contain unique information about the variable supposed to be the effect. It is the same idea proposed 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*'.

The question now is finding the formula to translate the concept of predictability to a single cross-section of contemporaneous data, without temporal perspective. We explore the possibilities of a non-parametric method for measuring causal information transfer between systems, called *transfer entropy* as proposed by Schreiber (2000). Marschinski and Kantz, (2002), or Dicks and Panchenko, (2006), work in the same direction.

In fact, our own proposal is based on *entropy* (see Joe 1989 a and b, Hong and White, 2005, and references from there), a flexible non-parametric technique aimed at finding regular patterns in large collections of data. Matilla and Ruiz (2008) introduce *symbolic dynamics*, with the purpose of summarizing fundamental pieces of information, and *symbolic entropy*, as a way of measuring the quantity of this information.

Section 2 sets the notation, definition and basic elements of our approach. In Section 3 we introduce the test of spatial causality, which is based on the comparison of two measures of conditional entropy. In Section 4 we discuss the problem of how a series can be symbolized. In Section 5the method is applied to the case of the relation between employment and migration, using a set of Italian data. Main conclusions appear in the sixth section.

2. Preliminaries

Let us assume a spatial process $\{x_s\}_{s \in S}$, where S is a set of points or locations in space.

We say that x is embedded in an m-dimensional space ($m \in N$ with $m \geq 2$) in reference to the following (mx1) vector:

$$\mathbf{x}_m(s_0) = \left\{x_{s_0}, x_{s_1}, \ldots, x_{m-1}\right\} \quad \text{for} \quad s_0 \in S \qquad [1]$$

where $s_1, s_2, \ldots, s_{m-1}$ are the m-1 nearest-neighbors to s_0, ordered according to distance with respect to location s_0. If some locations are equidistant to s_0, they are ordered counterclockwise.

Let $\Gamma_n = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ be a set of n symbols and assume that there is a map:

$$f : R^m \rightarrow \Gamma_n \qquad [2]$$

such that each element x_s is associated with a single symbol $f(x_s)=\sigma_{is}$ with $i_s \in \{1, 2, \ldots, n\}$. We say that location $s \in S$ is of the σ_i-type, in relation to $\{x_s\}_{s\in S}$, if and only if $f(\mathbf{x}_m(s)) = \sigma_i^x$. We call *f* the *symbolization map*.

Moreover, by $p(\sigma)$ we mean the probability of observing symbol σ. The set $\{p(\sigma_1);p(\sigma_2);...;p(\sigma_n)\}$ is the discrete probability function for the set of symbols, whose information content can be assessed with, for example, the Shannon (1948) entropy measure:

$$h = -\sum_{\sigma\in\Gamma_n} p_\sigma \ln(p_\sigma) \qquad [3]$$

The lower bound is attained if only one symbol occurs; the upper bound corresponds to the case where all the symbols appear with the same probability. It is conventionally assumed that 0 *log* 0 = 0.

This measure depends on a sequence of unknown probabilities, p_σ, that should be estimated from the data. Using the results of the symbolization map of [2], we may define the absolute frequency of a given symbol for the *x* data:

$$n_{\sigma_i^x} = \#\left\{s \in S \middle| s \text{ is } \sigma_i^x - \text{type for } x\right\} \qquad [4]$$

which coincides with the cardinality of the subset of *S* formed by all elements of σ_i-type. The relative frequency of a symbol, $\sigma_i^x \in \Gamma_n$, is a consistent estimate of its probability, for a great variety of cases:

$$p(\sigma_i^x) \equiv p_{\sigma_i^x} = \frac{\#\left\{s \in S \middle| f(\mathbf{x}_m(s)) = \sigma_i^x\right\}}{|S|} \qquad [5]$$

where |*S*| denotes the cardinality of set *S*. Using this (estimated) probabilities, we can measure, for a given embedding dimension $m \geq 2$, the Shannon entropy for the $\{x_s\}_{s\in S}$ process:

$$h_m(X) = -\sum_{\sigma_i^x\in\Gamma_n} p_{\sigma_i^x} \ln(p_{\sigma_i^x}) \qquad [6]$$

Consider now a bivariate process $\{\mathbf{Z}_s = (X_s, Y_s)\}_{s\in S}$. We define a new set of *joint* symbols, Ω_n, as the direct product of the two univariate sets $\Omega_n^2 = \Gamma_n x \Gamma_n$, with typical elements $\eta_{ij} = \left(\sigma_i^x; \sigma_j^y\right)$. The symbolization function is a mere generalization of the univariate case:

$$g(\mathbf{z}_s)_{s \in S} \to \Omega_n^2 \qquad [7]$$

Where:

$$g(\mathbf{z}_s = (X_s, Y_s)) = (f(x_s); f(y_s)) = \left(\sigma_i^x; \sigma_j^y\right) = \eta_{ij} \qquad [8]$$

The measure of *joint* entropy for the bivariate process is, formally, the same

$$h_m(z) = -\sum_{\eta \in \Omega_n^2} p_\eta \ln(p_\eta) \qquad [9]$$

We can define the conditional entropy of one variable, say *y*, conditional on the realization of a certain symbol in the other, for example σ^x in *x*:

$$h_m(y|\sigma^x) = -\sum_{\sigma^y \in \Gamma_n} p(\sigma^y|\sigma^x) \ln(p(\sigma^y|\sigma^x)) \qquad [10]$$

And also the conditional symbolic entropy of *y* given *x*

$$h_m(y|x) = -\sum_{\sigma^x \in \Gamma} \sum_{\sigma^y \in \Gamma_n} p(\sigma^y; \sigma^x) \ln(p(\sigma^y|\sigma^x)) \qquad [11]$$

It can be shown that:

$$h_m(y|x) = \sum_{\sigma^x \in \Gamma} p(\sigma^x) h_m(y|\sigma^x) \qquad [12]$$

The same as before, the different probabilities that appear in expressions [9]-[12] can be estimated by means of the relative frequencies:

$$p(\sigma_i^x) = \sigma_i^x / |S|;\ p(\sigma_j^y) = \sigma_j^y / |S|;\ p(\eta_{ij}) = n_{\eta_{ij}} / |S| \qquad [13]$$

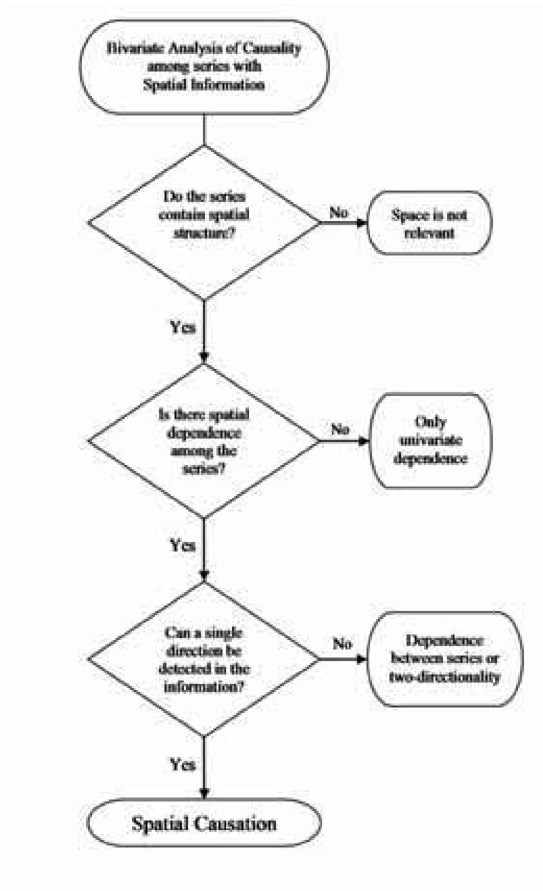
3. A Procedure for Testing Spatial Causality (in information)

As said, testing for causality is not simple. Difficulties increase in a spatial cross-section where, habitually, the relations are simultaneous. The absence of a previous ordering of the data is a serious constraint that limits our possibilities. The approach that follows is simple and is based on the information content provided by one variable, supposed to be the cause, with respect to another variable, supposed to be the effect, once all the information existing in the Space in relation to the 'caused' variable has been taken into account. That is, the first variable must contain unique information in relation to the second variable.

Previously, we have to make sure that (1)- Space is relevant to interpret the information of the two series and (2)- The two series are not independent. If both series are spatially independent, the Space is not relevant for them, so a traditional causality approach for cross-sections can be used. Besides, if the two series were independent, it would not make sense to think about causality. The proposal that follows assumes that both clauses are satisfied (space is relevant and the two series are related). Figure 1 shows the sequence of actions.

Figure 1

Testing for causality between spatial series



The Granger notion of causality builds upon the concept of *incremental predictability*, in the sense that:

$$P\left[y_{t+1} \middle| \Omega_t\right] \neq P\left[y_{t+1} \middle| \Omega_t - x_t\right] \qquad [14]$$

P[-] being a measure of probability, $\{\Omega_t\}$ refers to the knowledge available at time t, including the past history of the variable supposed to be 'caused', $y_t \in \Omega_t$, and $\{\Omega_t - y_t\}$ is the knowledge in the Universe available at time t except the information corresponding to variable x up to time t, $\{x_t\}$. However, the concept of predictability is not well developed yet for the case of pure cross-sections (in spite of, for example, Kelejian and Prucha, 2007). We prefer to think in terms of *incremental information*. The measures of entropy introduced in the last section offer us an adequate framework to progress in this direction.

In relation to the formalization of the space, we follow the usual reasoning in terms of sequences of weighting matrices. By *spatial dependence structure* we mean the set formed by all relevant weighting matrices that may intervene in a causal relationship between the two processes, x and y, $\mathbf{W}(\mathbf{x},\mathbf{y}) = \left\{\mathbf{W}_j \middle| j \in J\right\}$, J being a set of indices. $\mathbf{W}(\mathbf{x},\mathbf{y})$, contains a finite, and usually small, set of elements. Then, we can define the sets of spatially lagged information for both variables:

$$\begin{aligned} \mathbf{x_W} &= \left\{\mathbf{W}_j\mathbf{x} \middle| \mathbf{W}_j \in \mathbf{W}(\mathbf{x},\mathbf{y})\right\} \\ \mathbf{y_W} &= \left\{\mathbf{W}_j\mathbf{y} \middle| \mathbf{W}_j \in \mathbf{W}(\mathbf{x},\mathbf{y})\right\} \end{aligned} \qquad [15]$$

Finally, we propose the following definition of spatial causality, in information:

We say that $\{x_s\}_{s\in S}$ *does not cause* $\{y_s\}_{s\in S}$, *under the causal spatial structure* **W(x,y)**, *if*

$$h_m \mathbf{y} \middle| \mathbf{y}_W = h_m \mathbf{y} \middle| \mathbf{y}_W ; \mathbf{x}_W \qquad [16]$$

This definition leads us to the following non-parametric test for the null hypothesis of non-causality:

$$\left.\begin{array}{l} H_0: \begin{array}{l} \{x_s\}_{s\in S} \text{ does not cause } \{y_s\}_{s\in S}, \\ \text{under the causal spatial structure } \mathbf{x}_W \text{ and } \mathbf{y}_W \end{array} \\ H_A: \text{No } H_0 \end{array}\right\} \qquad [17]$$

and to the following statistics:

$$\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W) = \hat{h}_m(\mathbf{y}\middle|\mathbf{y}_W) - \hat{h}_m(\mathbf{y}\middle|\mathbf{y}_W;\mathbf{x}_W) \qquad [18]$$

Obviously, if $\mathbf{x}_W$ does not contain extra information about $\mathbf{y}$ then $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W) \simeq 0$, otherwise $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W) > 0$

The distribution function of the statistic $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)$ is unknown. In order to test the hypothesis (17) as an approximation, we propose bootstrapping the data under the null hypothesis of no causality. The procedure, with a number of *B* replicas, consists of the following steps:

1. Compute the value of the statistic $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)$ from the original sample.

2. Re-sampling $\{x_s\}_{s\in S}$ and $\{y_s\}_{s\in S}$ we obtain two bootstrapped series $\{x_s^b\}_{s\in S}$ and $\{y_s^b\}_{s\in S}$, where *b* is the number of bootstrapped samples.

3. For series $\{x_s^b\}_{s\in S}$ and $\{y_s^b\}_{s\in S}$, estimate the statistic: $\hat{\delta}^b(\mathbf{y}_W;\mathbf{x}_W)$.

4. Repeat *B*-1 times steps 2 and 3 to obtain *B*-1 bootstrapped realizations of the statistic $\left\{\hat{\delta}^b(\mathbf{y}_W;\mathbf{x}_W)\right\}_{b=1}^{B}$

5. Compute the estimated bootstrap *p-value*:

$$p_{boots} - \text{value}(\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)) = \frac{\sum_{b=1}^{B} \tau\left[\hat{\delta}^b(\mathbf{y}_W;\mathbf{x}_W) > \hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)\right]}{B}$$

Where $\tau[-]$ is an indicator function that takes the value 1 if the inequality inside the brackets is true and 0 otherwise.

6. For a significance level *α*, reject the null hypothesis that $\{x_s\}_{s\in S}$ does not cause $\{y_s\}_{s\in S}$ under the spatial structure $\mathbf{W}(\mathbf{x},\mathbf{y})$ if

$$p_{boots} - \text{value}(\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)) < \alpha$$

Herrera (2011) studies the behaviour of the $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)$ test under different circumstances through a large Monte Carlo. The conclusions of this study are promising in the sense that it enables us to determine (1)- the existence, or not, of causality between a pair of series and (2)- to identify, unequivocally, the direction of the information flow between series. For small sample sizes (100 observations) the test works better in the case of linear relationships between the variables; for large sample sizes (400 observations), the test is fairly robust to the functional form.

4. Symbolizing the series

This section is devoted to the problem of symbolizing a (spatial) series. The specification of a symbolization map is a prerequisite for apply the methodology presented in the last section, which allows us to solve the inference in the space of symbols instead of the sampling space. The decision is of paramount importance although in a given situation there will surely be various alternatives. The procedure shown below has proven useful for analyzing causality in pairs of spatial processes; however, it can be adapted and refined according to the requirements of the problem at hand.

Denote by *Me* the median of the spatial process $\{x_s\}_{s \in S}$ and define the indicator function:

$$\tau_s = \begin{cases} 0 & \text{if} \quad x_s \leq Me^x \\ 1 & \text{otherwise} \end{cases} \qquad [19]$$

As before, let $m \geq 2$be the embedding dimension that allows us to build the *m-surrounding* of each location in space. Let us remind that each *m-surrounding* contains the observation at point *s* plus the *m*-1 nearest-neighbors In continuation, the following indicator function, for each s_i with $i = 1, 2, \ldots, m\text{-}1$, is defined:

$$\iota_{s s_i} = \begin{cases} 0 & \text{if} \quad \tau_s \neq \tau_{s_i} \\ 1 & \text{otherwise} \end{cases} \qquad [20]$$

We are in a position of specifying a symbolization map for $\{x_s\}_{s \in S}$ simply as:

$$f(x_s) = \sum_{i=1}^{m-1} \iota_{s s_i} \qquad [21]$$

The set of symbols is $\Gamma_m = \{0;1;2,\ldots;m\text{-}1\}$ with cardinality equal to *m*. In sum, this symbolization process consists in comparing, for each location *s*, the value of the indicator function (19), τ_s, with the same indicator for the points included in the *m*-surrounding of location *s*, τ_{s_i}. The symbols measure the number of coincidences.

In continuation we present two examples that show the use of the above methodology to test for spatial causality between two spatial variables. The sample size is 100 and the coordinates of the observation points, abscissa and ordinate, are randomly distributed in the unit interval.

The first example corresponds to the case where one variable, **y**, is caused in a unidirectional way by another variable, **x**. We have used the following equations to obtain the data:

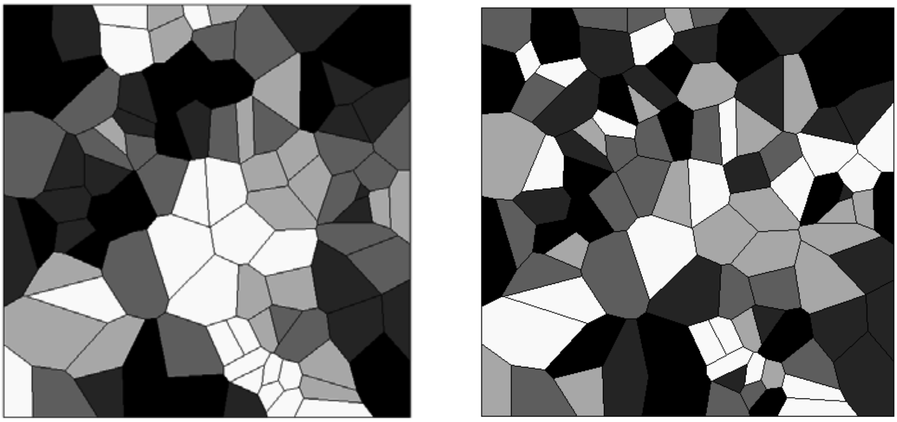
$$\left.\begin{array}{l}\mathbf{y}=(\mathrm{I}-\rho\mathbf{W})^{-1}\left(\beta_0\mathrm{x}+\beta_1\mathbf{Wx}+\mathbf{u}\right)\\ \mathbf{x}\sim\mathrm{iidN}(0,1);\ \mathbf{u}\sim\mathrm{iidN}(0,1)\\ \mathrm{Cov}(\mathbf{x},\mathbf{u})=0\end{array}\right\} \qquad [22]$$

where ρ=0.5; β_0=1 and β_1=2 (different values produce similar results). The values obtained for **y** and **x** are represented in the 5 quintiles of Figure 2 (darker color, higher quintile).

Figure 2

Example 1. There exists a causality relation from x to y

Spatial distribution of **y** Spatial distribution of **x**



The $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)$ statistic (18) is a difference between two measures of conditional entropy. The same as with probabilities, these **conditional** entropies can be obtained as the difference between the corresponding joint and the marginal entropies.

$$\begin{array}{l}\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)=\hat{h}_m(\mathbf{y}|\mathbf{y}_W)-\hat{h}_m(\mathbf{y}|\mathbf{y}_W;\mathbf{x}_W)=\\ \left[\hat{h}_m(\mathbf{y};\mathbf{y}_W)-\hat{h}_m(\mathbf{y}_W)\right]-\left[\hat{h}_m(\mathbf{y};\mathbf{y}_W;\mathbf{x}_W)-\hat{h}_m(\mathbf{y}_W;\mathbf{x}_W)\right]\end{array} \qquad [23]$$

For the case m=4, there are 4 symbols to symbolize one variable, $\Gamma_4=\{0,1,2,3\}$, 16 if two series are symbolized $\Omega_4^2=\Gamma_4\times\Gamma_4=\{(0.0);(0.1);...(3.2),(3.3)\}$ and 64 in the case of symbolizing three variables $\Omega_4^3=\Gamma_4\times\Gamma_4\times\Gamma_4=\{(0,0,0);(0,0,1);...(3,3,2),(3,3,3)\}$. Table 1 shows some of the data included in Figure 2 and, as a mere example, the symbolization of two series, **y** and $\mathbf{y}_W$, and the obtaining of the bivariate symbols, in the last column of the Table.

Table 1

EXAMPLE 1. Symbolization process for two series afected by a causality relationship

(Continues)

Coordinates		Data		3-surroundings				$y_{(S_i)}$				$y_{w(S_i)}$				
Nob	yc	xc	y	x	s_0	s_1	s_2	s_3	$y_{(S_0)}$	$y_{(S_1)}$	$y_{(S_2)}$	$y_{(S_3)}$	$y_{w(S_0)}$	$y_{w(S_1)}$	$y_{w(S_2)}$	$y_{w(S_3)}$
1	0.98	0.64	1.85	0.33	**1**	62	18	91	1.85	1.49	1.82	1.97	1.76	1.88	1.77	1.65
2	0.56	0.18	-1.34	-1.77	**2**	74	54	67	-1.34	-1.20	0.55	-0.59	-0.42	-0.01	-0.35	-0.66
3	0.03	0.73	0.92	-2.27	**3**	60	76	29	0.92	1.14	-0.12	-0.71	0.10	0.29	0.47	1.00
4	0.36	0.91	0.55	-1.32	**4**	25	66	8	0.55	-0.28	-0.09	0.30	-0.03	-0.19	0.32	-0.39
5	0.29	0.68	-1.80	0.47	**5**	22	28	43	-1.80	0.85	1.04	0.01	0.63	-0.39	-0.31	0.03
….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….
95	0.91	0.27	1.06	0.99	**95**	37	83	27	1.06	1.49	0.11	0.16	0.59	1.03	1.18	0.89
96	0.13	0.12	-0.12	-0.81	**96**	92	57	73	-0.12	0.92	-0.24	-0.66	0.01	-0.34	-0.02	0.17
97	0.87	0.92	0.07	1.56	**97**	10	85	12	0.07	0.68	0.11	2.10	0.96	0.82	0.29	-0.12
98	0.97	0.05	-0.36	-0.07	**98**	21	27	95	-0.36	-0.61	0.16	1.06	0.20	0.29	0.89	0.59
99	0.41	0.02	-1.18	0.80	**99**	64	89	56	-1.18	0.51	-1.03	-0.44	-0.32	-0.88	-0.76	0.36
100	0.64	0.94	-1.71	0.75	**100**	88	93	20	-1.71	0.92	-1.34	0.28	-0.05	-0.31	1.10	0.50

(Conclusion)

	$y_{(S_i)}$				$y_{w(S_i)}$				$y_{(S_i)}$			$y_{w(S_i)}$			η_{ij}	
Nob	τ_{so}	τ_{s1}	τ_{s2}	τ_{s3}	τ_{so}	τ_{s1}	τ_{s2}	τ_{s3}	$\iota_{so\ s1}$	$\iota_{so\ s2}$	$\iota_{so\ s3}$	$\iota_{so\ s1}$	$\iota_{so\ s2}$	$\iota_{so\ s3}$	σ_i^x	σ_j^y
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**3**	**3**
2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	**3**	**3**
3	0	0	1	1	0	1	1	1	1	0	0	0	0	0	**1**	**0**
4	0	0	1	0	0	0	1	0	1	0	1	1	0	1	**2**	**2**
5	1	0	0	0	1	0	0	0	0	0	0	0	0	0	**0**	**0**
….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….
95	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**3**	**3**
96	0	0	0	0	0	0	0	0	1	1	1	1	1	1	**3**	**3**
97	1	1	0	0	1	1	1	0	1	0	0	1	1	0	**1**	**2**
98	0	0	1	1	0	1	1	1	1	0	0	0	0	0	**1**	**0**
99	0	0	0	1	0	0	0	1	1	1	0	1	1	0	**2**	**2**
100	0	0	1	1	0	0	1	1	1	0	0	1	0	0	**1**	**1**

This information allows us to obtain the estimated probabilities for the univariate symbols of **y** and **x**:

	p(0)	p(1)	p(2)	p(3)
y	0.03	0.14	0.23	0.60
x	0.10	0.26	0.34	0.30

The probabilities corresponding to bivariate symbols are simple to obtain:

	p(0,0)	p(1,0)	p(2,0)	p(3,0)	p(0,1)	p(1,1)	p(2,1)	p(3,1)	p(0,2)	p(1,2)	p(2,2)	p(3,2)	p(0,3)	p(1,3)	p(2,3)	p(3,3)
$\mathbf{y},\mathbf{y}_W$	0.02	0.00	0.00	0.01	0.00	0.05	0.07	0.02	0.01	0.09	0.08	0.05	0.00	0.04	0.08	0.48
$\mathbf{x},\mathbf{x}_W$	0.00	0.08	0.02	0.00	0.03	0.12	0.10	0.01	0.04	0.15	0.13	0.02	0.03	0.08	0.15	0.04

as well as the probabilities for the 64 symbols of the trivariate case, $(\mathbf{y};\mathbf{y}_w;\mathbf{x}_w)$. The results that we obtain for this example are:

$$\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W) = \left[\hat{h}_m(\mathbf{y};\mathbf{y}_W) - \hat{h}_m(\mathbf{y}_W)\right] - \left[\hat{h}_m(\mathbf{y};\mathbf{y}_W;\mathbf{x}_W) - \hat{h}_m(\mathbf{y}_W;\mathbf{x}_W)\right]$$
$$= [1.836 - 1.025] - [2.597 - 2.229] = 0.443$$

$$\hat{\delta}(\mathbf{x}_W;\mathbf{y}_W) = \left[\hat{h}_m(\mathbf{x};\mathbf{x}_W) - \hat{h}_m(\mathbf{x}_W)\right] - \left[\hat{h}_m(\mathbf{x};\mathbf{x}_W;\mathbf{y}_W) - \hat{h}_m(\mathbf{x}_W;\mathbf{y}_W)\right]$$
$$= [2.394 - 1.308] - [3.156 - 2.229] = 0.160$$

The p-value of the first statistic, $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)$, after a random permutation, is 0.02 whereas that of the second, $\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W)$, is 0.31 which allows to state that variable **x** causes (in information) variable **y** under the spatial structure defined by the 3 nearest-neighbors criterion and using a spatial surrounding of m=4.

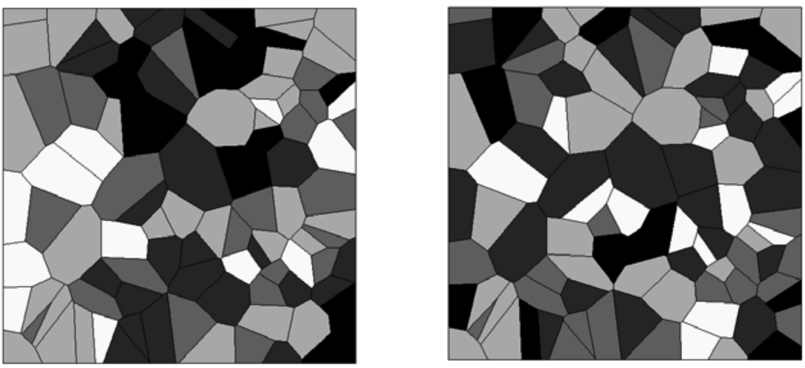
Figure 3 presents the case of two independently generated series:

$$\left.\begin{array}{l} y \sim \text{iidN}(0,1) \\ x \sim \text{iidN}(0,1) \\ \text{Cov}(x,y) = 0 \end{array}\right\} \quad [24]$$

Figure 3

Example 2. There is no causal relationship between x and y

Spatial distribution of **y** Spatial distribution of **x**



The estimated probabilities for the symbols of **y** and **x** are:

	p(0)	p(1)	p(2)	p(3)
y	0.11	0.23	0.28	0.38
x	0.09	0.25	0.39	0.27

And the probabilities corresponding to the bivariate symbols are:

	p(0,0)	p(1,0)	p(2,0)	p(3,0)	p(0,1)	p(1,1)	p(2,1)	p(3,1)	p(0,2)	p(1,2)	p(2,2)	p(3,2)	p(0,3)	p(1,3)	p(2,3)	p(3,3)
$\mathbf{y},\mathbf{y}_W$	0.02	0.04	0.05	0.00	0.00	0.13	0.09	0.01	0.02	0.12	0.12	0.02	0.06	0.05	0.18	0.09
$\mathbf{x},\mathbf{x}_W$	0.01	0.07	0.01	0.00	0.04	0.10	0.09	0.02	0.05	0.22	0.11	0.01	0.01	0.07	0.12	0.07

From the measures of probability we obtain:

$$\hat{\delta}(\mathbf{y}_W;\mathbf{x}_W) = \left[\hat{h}_m(\mathbf{y};\mathbf{y}_W) - \hat{h}_m(\mathbf{y}_W)\right] - \left[\hat{h}_m(\mathbf{y};\mathbf{y}_W;\mathbf{x}_W) - \hat{h}_m(\mathbf{y}_W;\mathbf{x}_W)\right]$$
$$= [2.394 - 1.305] - [3.436 - 2.552] = 0.205$$

$$\hat{\delta}(\mathbf{x}_W;\mathbf{y}_W) = \left[\hat{h}_m(\mathbf{x};\mathbf{x}_W) - \hat{h}_m(\mathbf{x}_W)\right] - \left[\hat{h}_m(\mathbf{x};\mathbf{x}_W;\mathbf{y}_W) - \hat{h}_m(\mathbf{x}_W;\mathbf{y}_W)\right]$$
$$= [2.377 - 1.284] - [3.406 - 2.552] = 0.239$$

Clearly the two are not significant. The p-values are 0.185 and 0.221.

Before continuing, we must recognize that, in the process of symbolization, some information will be lost. The symbols offer a coarse description of the data, they are very flexible and the inference will be, in general, easier in the symbols space than in the sampling space. However, it is clear that they will retain only a part of the sampling characteristics. The idea it is that the user must make a judicious choice in order to assure that the information maintained after the symbolization will be the most important in relation to the hypothesis that we are testing (Lopez et al, 2010).

5. Migration vs. unemployment. An application to the Italian case

This part of the paper applies our methodology to a real case, the relation between migration and unemployment using data for the Italian provinces over the period 1996-2005.

The characteristics of this relation are controversial in the literature. Many colleagues argue that immigration is the cause of high unemployment in regions receiving large number of migrants. Assuming, from a neoclassical perspective, that labor is homogeneous and there is perfect competition on the goods market, workers move to prosperous regions, increasing the labor supply there (this is the direct effect). In turn, immigrants increase the consumption of local goods, improving labor demand (this is the indirect effect). According to the neoclassical perspective, the direct effect prevails over the indirect effect, resulting in an increase in unemployment.

New Economic (Krugman, 1991) also supports the existence of causal relationships between migration and unemployment. Assuming imperfect competition on the goods market and rigidity on the labor market, Epifani and Gancia (2005) show that the forces that generate agglomeration also determine the spatial disparities in unemployment. Regional integration results from diminishing transport costs, which stimulate migration to prosperous regions. The migration flows create agglomeration economies (home-market effect), more activity and, therefore, an increasing demand for labor. In this case, the indirect effects of immigration on labor demand will prevail over the direct effect. The core-periphery dichotomy is being reinforced, with immigration reducing the unemployment rate in the region of destiny.

Other colleagues suggest that, on the contrary, unemployment is the cause of migration. Pissarides and Wadsworth (1989) argue that people move from places where they are not employed to places offering greater possibilities to get a job. Unemployment in the place of origin also increases the probability of migration, as people are more likely to become unemployed or continue in unemployment.

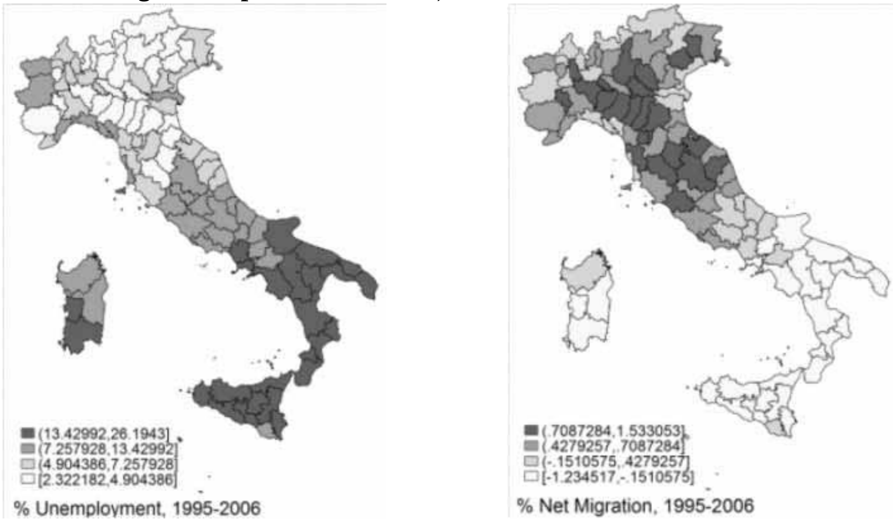
In sum, there are important arguments supporting both directions of causality. This is an interesting study case in the literature, especially for our approach to spatial causality (see Herrera, 2001, for a more thorough discussion).

We use annual data for the NUTS 3 regions obtained from the Italian National Statistics Institute (ISTAT) for the period 1995-2006. The unemployment rate is unemployed divided by labor force. The net migration rate is the average net migratory balance, new migrants minus the number of migrants leaving municipal censuses, divided by the total population (aged 15-64). In sum, we have the average for the period 1995-2006 for both variables and for the 103 Italian provinces. The situation is shown in Figure 4.

The average unemployment rate during the period was 9.28% in a clearly declining process, from as 10.73% at the beginning of the period to a 7.59% over the last years. The average migration is 0.55% with a rising profile (which doubled the rate of the first years, 0.27% to 0.64% at the end). The spatial distribution of the data is very characteristic, with a huge concentration of unemployment in the southern provinces which also show a negative migratory balance. This picture has remained the same, with only slight variations over the last decades.

Figure 4

Migration and Unemployment in the Italian provinces (annual average of the period 1995-2006)



We use a row-standardized weighting matrix build on the three nearest-neighbors criterion, which allows us to confirm that both variables are not randomly distributed over the Italian provinces and that, indeed, they are dependent. In Table 2 we include the Moran's I for the univariate test of spatial dependence (Moran, 1950) and the bivariate Moran's I (Czaplewski and Reich, 1993) for the test of spatial independence between the variables of unemployment and migration.

Table 2

Unemployment vs Migration. The Italian case (1995-2006)

	Unemployment	Migration
Univariate Moran' I	0.432 (0.000)	0.661 (0.000)
Bivariate Moran's I	-0.780 (0.000)	
CAUSALITY ANALYSIS	*Unemp.* $\not\Rightarrow$ *Migra.*	*Migra.* $\not\Rightarrow$ *Unemp.*
	0.042 (0.681)	0.153 (0.000)

NOTE: pvalues in parenthesis. The symbol "$\not\Rightarrow$" means "no-causes".

According to the results shown in Table 2, there is a clear causal relationship (in information), for the case of the Italian data by provinces, from net migration to unemployment. The conclusion is clear and, according to Herrera (2011), this relation has been fairly stable for the different sub-periods.

6. Conclusions and further research

To our knowledge, this is the first article, after Blommestein and Nijkamp (1981), which explicitly brings to the fore the problem of testing for causality between spatial variables, as a first step in the specification of a spatial econometric model. The next reference is Herrera (2011) to which we are clearly indebted.

Our intention was to develop an empirically testable notion of spatial causality, using the generally accepted concept of incremental informative content. Intuitively, our definition establishes that causality implies that the *cause* variable should provide additional information about the *effect* variable. The key concept is the term *information* in the sense of *"numerical quantity that captures the uncertainty in the result of an experiment to be conducted"*. This definition makes direct reference to the measures of entropy of an information set. In this sense, the test of causality presented in this paper compares two measures of conditional entropy. The first uses as conditional variables all the information that exists in the space concerning the variable supposed to be the effect; the second measure uses the same set of conditioning variables but the variable supposed to be the cause. The test is intuitive, simple to obtain and does not need any hypothesis about functional form, distribution function, or other aspects of the specification. It is a fully non-parametric causality test.

Future research will enhance the analysis by including more than two variables, in order to approach the common cause principle. Another strand of future research points towards the inclusion of time in this framework, in order to deal with spatio-temporal or panel data.

Acknowledgements: The authors would like to express their thanks to the project ECO2009-10534 of the *Ministerio de Ciencia e Innovación del Reino de España*.

REFERENCIAS

BAUWENS, L., BOSWIJK, P. AND URBAIN, P. (2006). «Causality and Exogeneity in Econometrics». *Journal of Econometrics*, 102, 305–309.

BLOMMESTEIN, H. AND NIJKAMP, P. (1981). «Soft Spatial Econometric Causality Models». *Papers in Regional Science*, 51, 65–77

CZAPLEWSKI, R. AND REICH, R. (1993). «Expected Value and Variance of Moran's Bivariate Spatial Autocorrelation Statistic under Permutation». *USDA For. Serv. Res.* Paper RM-309, Fort Collins, Colorado, 1–13.

DIKS, C. AND PANCHENKO, V. (2006). «A New Statistic and Practical Guidelines for Nonparametric Granger Causality Testing». *Journal of Economic Dynamics and Control*, 30, 1647–1669.

EPIFANI, P. AND GANCIA, G. (2005). «Trade, Migration and Regional Unemployment». *Regional Science and Urban Economics*, 35, 625–644.

GRANGER, C. (1980). «Testing for Causality: A Personal Viewpoint». *Journal of Economic Dynamics and Control*, 2, 329–352.

HAAVELMO, T. (1944). «The Probability Approach in Econometrics». *Econometrica*, 12 Supplement, iii-vi+1–115.

HERRERA, M. (2011). *Causality. Contributions to Spatial Econometrics*. Ph.D. Thesis, University of Zaragoza. Available in https://sites.google.com/site/spatialcausality/.

HONG, Y. AND WHITE, H. (2005). «Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence». *Econometrica*, 73 837–901.

HOOVER, K. (2004). «Lost causes». *Journal of the History of Economic Thought*, 26, 149–164.

HOOD III, M., KIDD, Q. AND MORRIS, I. (2008). «Two Sides of the Same Coin? Employing Granger Causality Tests in a Time Series Cross-Section Framework». *Political Analysis*, 16, 324–344.

HURLIN, C. AND VENET, B. (2001). «Granger Causality Tests in Panel Data Models with Fixed Coefficients». University of Paris Dauphine: Working Paper Eurisco 2001–09.

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 PRUCHA, I. (2007). «The Relative Efficiencies of Various Predictors in Spatial Econometric Models Containing Spatial Lags». *Regional Science and Urban Economics*, 37, 363–374.

KRUGMAN, P. (1991). «Increasing Returns and Economic Geography». *Journal of Political Economy*, 99 483–499.

LÓPEZ, F. ,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.

MARSCHINKI, R. AND KANTZ, H. (2002). «Analysing the Information Flow Between Financial Time Series. An Improved Estimator for Transfer Entropy». *The European Physical Journal B,* 30, 275–281.

MATILLA, M. AND RUIZ, M. (2008). «A Non-Parametric Independence Test Using Permutation Entropy». *Journal of Econometrics*, 144, 139–155.

MORAN, P. (1950). «A Test for the Serial Independence of Residuals». *Biometrika*, 37, 178–181.

MORGAN, M. (1991). *The History of Econometric Ideas.* Cambridge University Press, Cambridge.

PEARL, J. (2009). «Causal Inference in Statistics: An Overview». *Statistics Surveys*, 3, 96–146.

PISSARIDES, C. AND WADSWORTH, J. (1989). «Unemployment and the Interregional Mobility of Labour». *Economic Journal*, 99, 739–755.

SCHREIBER, T. (2000). «Measuring Information Transfer». *Physical Review Letters*, 85, 461–464.

SHANNON, C. (1948). «A mathematical theory of communication». *Bell System Technical Journal*, 27, 379–423. 623-656, July, October.

TERVO, H. (2009). «Centres and Peripheries in Finland: Granger Causality Tests Using Panel Data». *Spatial Economic Analysis*, 4, 377–390.

TINBERGEN, J. (1951). *Econometrics*. Palgrave Macmillan, London.

WEINHOLD, D. AND NAIR, U. (2001). «Causality Tests for Cross Country Panels: A New Look at FDI and Economic Growth in Less Developed Countries». *Oxford Bulletin of Economics and Statistics*, 63, 153–72.

WIENER, N. (1956). «The Theory of Prediction». In Beckenbach, E. (ed.), *Modern Mathematics for Engineers*. McGraw, New York.