doi:10.1111/pirs.12144

Detecting causal relationships between spatial processes*

Marcos Herrera[1], **Jesús Mur**[2], **Manuel Ruiz**[3]

[1] CONICET-IELDE, National University of Salta, Salta, 4400, Argentina (e-mail: mherreragomez@gmail.com)
[2] Department of Economic Analysis, University of Zaragoza, Zaragoza, 50005, Spain (e-mail: jmur@unizar.es)
[3] Department of Computing and Quantitative Methods, Technical University of Cartagena, Cartagena, 30203, Spain (e-mail: manuel.ruiz@upct.es)

Received: 5 September 2013 / Accepted: 6 August 2014

Abstract. The paper focuses on the problem of testing for the existence and direction of causality in a group of variables, with a spatial framework. Our objective is to produce useful criteria to discuss causality on rational and objective bases. We introduce a new non-parametric test, based on symbolic entropy, which is robust to the functional form of the relation. The test has good behaviour in samples of medium to large size. The proposal is illustrated with an application to the case of migration versus unemployment, using data on 3,108 US counties for the period 2003–2008.

JEL classification: C21, C50, R15

Key words: Non-parametric methods, spatial causality, spatial bootstrap, symbolic dynamics

1 Introduction

In recent years there has been a growing interest in causality issues, as shown by the bibliometric study of Hoover (2004, p.4): 70 per cent of the articles in the *JSTOR* archives, published during the last decades, contains words in a ‘causal family (“cause”, “causes”, “causal”, “causally” or “causality”)’. The percentage increases up to 80 per cent if the search is restricted to the field of econometric papers.

However, the impact of this debate has been surprisingly small in the field of spatial econometrics. For example, the subject index of the most influential textbook in this discipline, Anselin (1988), comprises approximately 500 items, not one of them being related to causation.

* We appreciate the constructive comments received from many colleagues, especially Jeff Racine, Jean Paelinck, Roberto Basile, Mariano Matilla and Fernando López. We also thank J. Paul Elhorst and two anonymous referees of *Papers in Regional Science* who provided numerous helpful comments which improved the paper. Marcos Herrera and Jesús Mur sincerely acknowledge the financial support received from the Spanish Government’s Ministry of Economy and Competitiveness (ECO2012-36032-C03-01). Jesús Mur and Manuel Ruiz wish to thank the support of COST Action IS1104 ‘The EU in the new economic complex geography: models, tools and policy evaluation’. Jesús Mur also thanks the Department of Industry and Innovation of the the Regional Government of Aragon for the research support. Manuel Ruiz was partially supported by MINECO (Ministerio de Economía y Competitividad) and FEDER (Fondo Europeo de Desarrollo Regional) projects ECO2012-36032-C03-03 and MTM2012-35240. The usual disclaimers apply.

The same can be said with respect to other popular textbooks, including the most recent of LeSage and Pace (2009): 1,000 headwords, none of which is related to Hoover's causal family.

There are several reasons to account for this singularity. First is the type of data, usually cross-sections without temporal perspective which clashes with the notion of a delay between cause and effect. Second is the multidirectionality of spatial interaction; in space there is nothing like an 'arrow of time' to interpret the data, which results in isotropic relationships. Moreover, it is assumed that a cross-sectional model is a general equilibrium model (Isard 1971) that must be interpreted, preferably, in the long-term where causality plays a minor role. This view is shared by part of the time series literature (i.e., Brockwell and Davis 2002).

In spite of these difficulties, we think that the notion of causality must be in the forefront of spatial econometrics methods. The widespread use of impact analysis illustrates this necessity. It is tempting to wonder, as LeSage and Pace (2009, p. 35) if 'a change in the explanatory variable for a single region (observation) can potentially affect the dependent variable in . . . other observations (regions)'. The statement has a strong causal meaning that can be answered only if we have a causal relation. Unfortunately this is not the case in most spatial models, which are essentially correlation models.

According to Pearl (2009, p. 138), the crucial distinction between the two lies in the interpretation of the equality sign: 'in structural (causal) models, this equality sign conveys the asymmetrical relation "is determined by" and hence behaves more like an assignment symbol (:=) . . . than like an algebraic equality'. For example, in the well-known model specified for the case of the crime in Columbus (Anselin 1988), the following three equations are small variations of each other:

$$y_i = \alpha_0 + \alpha_1 i_i + \alpha_2 h_i + \alpha_3 \mathbf{w}_i' y + u_i, \tag{1}$$

$$h_i = \beta_0 + \beta_1 i_i + \beta_2 y_i + \beta_3 \mathbf{w}_i' h + u_i, \tag{2}$$

$$i_i = \theta_0 + \theta_1 h_i + \theta_2 y_i + \theta_3 \mathbf{w}_i' i + u_i, \tag{3}$$

where y_i is an indicator of crime in the location i, i_i is income *per capita* and h_i housing value in the same location; the term $\mathbf{w}_i$ is a vector that captures 'nearby' locations. The first is the crime equation as appears in the textbook of Anselin. What makes singular this equation is the quality of asymmetry: there is certain spatial pattern in the variable but crime in the neighbourhood is determined (it is caused) by income and housing prices. If this is the correct conclusion, renders meaningless the Equations (2) and (3) because a variable cannot be, simultaneously, cause and effect.

However, if the variables proceed from a trivariate normal distribution, for example, the three system equation is admissible. In this case, parameter α_1 in (1) measures the reduction in variance of y_i obtained after conditioning for i_i. This is a totally different question from the impact on $E[y_i]$ due to an intervention in i_i. In fact, the answer can be zero if (1) is just a correlation equation (the same as the entire chain of direct, indirect and total effects).

Cartwright (1995) recognizes that causality cannot be derived only from functional relationships. Something additional is needed, which may be prior knowledge of the causal relation, as a prerequisite for any causal conclusion, or, as it is defended in this paper, statistical evidence supporting the assumption of causality and the asymmetry of the relation.

We believe this problem is of interest for applied research in spatial economics. The problem is how to address it. Angrist and Pischke (2009) propose performing natural experiments in economics. The aim is to mimic the conditions of an experiment where we have two groups (i.e., treatment and control) before and after the intervention. Gibbons and Overman (2012) argue that the experimentalist approach is a good alternative to achieve identification (the idea dates back to Isserman and Merrifield 1982), putting causality at the centre in the debate of spatial econometrics. However, the difficulty of this approach in a non-experimental field, as economics, is evident based on the uni-directionality of time.

The dominant paradigm in economics is non-experimentalist, favouring approaches such as that of Granger (1969) to the notion of causality. Unidirectionality of time, temporal precedence and information content are some of the main principles of his approach, supported also by Wiener (1956), as quoted by Hlaváčková-Schindler (2009, p. 185): ‘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’. This is the way that we want to explore in our contribution.

Recently, some researchers give the impression that spatial causality means: ‘when outcomes in area A affect outcomes in area B’ (Partridge et al. 2012, p. 168). For example, Gibbons and Overman (2012) offer an extended definition of spatial causality, using a spatial autoregressive model (*SAR*, in what follows):

$$y_i = \rho \mathbf{w}_i' y + \beta x_i + u_i, \tag{4}$$

where i indexes locations, y_i is the outcome, x_i an explanatory variable, u_i an error term, ρ and β are parameters. According to the authors, parameter β captures the causal effect of x on y and ρ represents the causal effect of neighboring dependent variable. This interpretation is firmly recorded in the DNA of spatial econometrics but, at the very least, it is questionable. Space is just the support where the data are taken. Indeed, we need to account of the stochastic structure of the variables (and spatial dependence is probably one of the most important) in order to isolate potential causality relations. However, space is not a causal factor the same that a variable cannot be a causal factor for itself. As indicated by Davidson (2000, p. 75): ‘Granger non-causality is a property of the DGP, under which (the information about x) can innocuously be excluded from the conditioning variables . . . It is unambiguously a relationship between variables’. According to this approach, the spatial lag in (4), $\mathbf{w}_i' y$, captures correlation but not causality. In fact, only x can cause y.

The objective of the paper is to introduce a new test for causality in pure cross-sectional spatial series, inspired on the Granger-Wiener’s idea of information content. Our proposal is fully non-parametric. The paper is structured as follows. Section 2 provides a review of some of the main principles of Granger causality. Section 3 introduces the machinery of our approach. Section 4 formalizes our causality test for spatial cross-sectional data. Section 5 contains the results of a Monte Carlo experiment, while Section 6 presents a case study, the relation between unemployment rate and net migration for US county data. Main conclusions appear in Section 7.

2 From Granger causality to spatial causality

As said, the Granger approach, together with the test of Sims (1972), is by far the preferred option to test for causality in applied econometrics. This literature has produced an operational and testable definition of causality, based on three principles:

1. Temporal precedence, in the sense that the effect should not precede the cause.
2. Temporal invariance that assures that causal mechanisms remain constant throughout time.
3. Information completeness, in the sense that all information needed for the variables involved in the analysis is available.

Under these premises, and assuming that the series are weakly stationary, we say according to Granger (1969) that x is a causal variable for y if and only if $V(y_{t+1} \mid \Lambda_t) < V(y_{t+1} \mid \Lambda_t')$. Λ_t refers

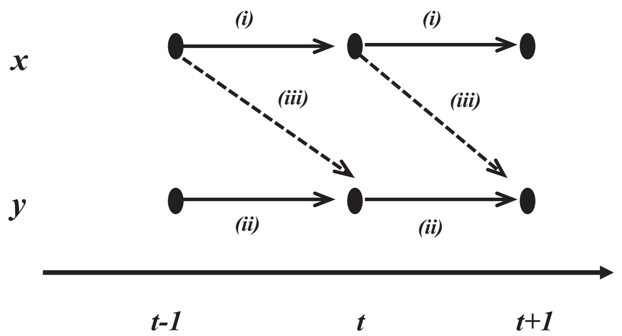

Fig. 1. Representation of Granger causality in a time domain

to the information set at time t, which includes historical information up to period t of both series, x_t and y_t, together with the contextual variables z_t. Moreover $\Lambda_t' \equiv \{y_{t-j}, z_{t-j}\}$, $\forall j \geq 0$, excludes the information of x_t and $V(\cdot)$ is the forecasting error variance. In brief, x causes y if future values of y can be predicted more precisely if past values of x, in relation to y_{t+1}, are included in the information set.

Let us emphasize a couple of important aspects of this definition:

1. Granger causality does not require the specification of a direct causal mechanism (that is, we do not need a model). Theory is important but this is not a theory-driven approach as, for example, Zellner (1988). All that matters is observable predictive capacity (Davidson 2000).
2. Investigators can never be sure whether an information set is complete. This is the problem of confounders, or common causes, which led Pearl (2009) to distinguish between necessary and sufficient conditions of causality: Granger causality is a necessary clause, not a sufficient one.

The functioning of the Granger test for a bivariate system in a time domain can be represented graphically as in Figure 1.

Relations (i) and (ii) reflect univariate serial dependence, relation (iii) adds causality from x to y. This means that if we want good predictions of y, it is not enough with the past of this series, we also need to control for the impact that x has on y. The principles of temporal precedence and time invariance appear clearly in this temporal flow-chart, which is naturally oriented. The third criterion, information completeness, refers to the order of the system and it is an uncertainty factor.

The spatial domain lacks of such natural orientation and relationships become more complicated because of the multidimensionality, as in the case of Figure 2 with three regions. As before, solid lines (relations (i) and (ii)) indicate relations of spatial dependence that affect, separately, each variable, x and y. This univariate spatial structure means spatial dependence, for x and y. If the spatial structure of both variables is similar, or there are unknown common factors shaping their spatial pattern, we will find spatial correlation between the two. However, this is not causality which is only reflected by the broken lines of (iii). The origin of the causal impulse may be in the same region (x causes y inside each region, and the spatial dependence of y boosts this relation outwards) or it may proceed from nearby regions, whose impact is transferred to the region because of the spatial dependence of y. As it is well-known, there are mirroring effects on space that complicate the analysis of spatial relationships.

The differences between both figures are evident. They illustrate that our proposal of detecting causality in cross-sections of spatial data will require specific techniques as those introduced in Section 3.

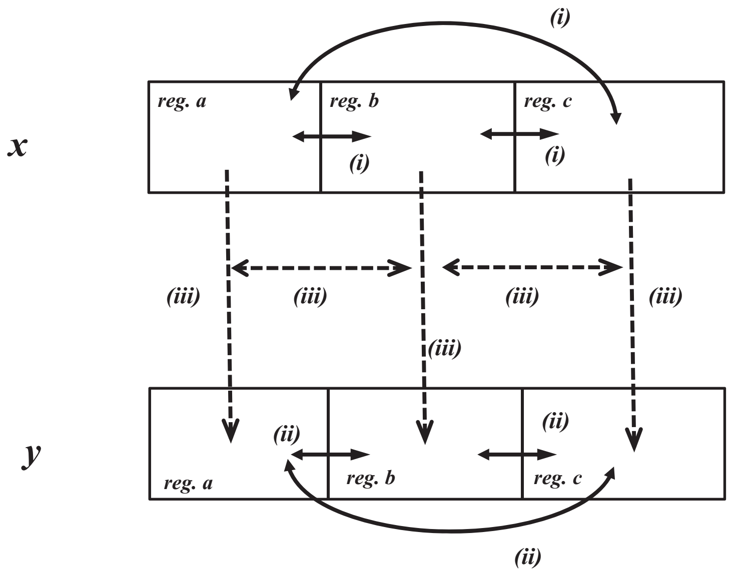

Fig. 2. Representation of Granger causality in a spatial domain

3 Symbolic dynamics and entropy

Symbolic dynamics appeared in the field of mathematics with the purpose of modelling dynamic systems by means of discrete sequences of (abstract) symbols obtained for a suitable partition of the state space (Hao and Zheng 1998). The basic idea is to consider a space in which the possible states of a system are represented; each state corresponds to a unique point in the state space. Then, this space can then be partitioned into a finite number of regions and each region can be labelled by a symbol (i.e., a Greek letter, a natural number). The inference is solved in the space of symbols. Symbolic dynamics is a coarse-grained method that loses some information, but retains most part of the essential features of the system.

The first step in symbolic dynamics is the so-called symbolization process, where the symbols are defined and assigned to the estate space. There is no a unique process of symbolization; in fact, we may think in different procedures among which the user must choose the most adequate. The observed distribution of the symbols conveys valuable information about the system that produced the data. This information can be summarized in a measure of symbolic entropy. Last step is solving the inference using this measure of information.

In the following we use symbolic entropy as a measure of spatial organization of a map. The idea is simple: if x causes y, information about the spatial distribution of x will help to significantly reduce the uncertainty about the distribution of y over space. First, the symbolization procedure is defined and then several measures of entropy, useful for our purpose, are introduced.

3.1 Symbolization process

Let us begin by defining two real spatial processes $\{x_s\}_{s\in S}$ and $\{y_s\}_{s\in S}$, where S is a set of geographical co-ordinates. The locations are given and fixed. Denote by $\Gamma_n = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ the non-empty finite set of n symbols. Symbolizing a process is defining a map

$$f : \{x_s\}_{s\in S} \to \Gamma_n, \tag{5}$$

such that each element x_s is associated to a unique symbol $f(x_s) = \sigma_i$. We say that location $s \in S$ is $\sigma_i - type$, relative to the series $\{x_s\}_{s\in S}$, if and only if $f(x_s) = \sigma_i$. We call f the 'symbolization map'. The same can be done for y.

Let us introduce the bivariate process $Z_s = \{x_s, y_s\}_{s\in S}$ and the corresponding set of symbols Ω_n, which may consists of the direct product of the two previous sets Γ_n. Formally, the symbolization map, g, for this Z is:

$$g:\{Z_s\}_{s\in S} \to \Omega_n = \Gamma_n \times \Gamma_n. \tag{6}$$

The elements of Ω_n are $\eta_{ij} = \left(\sigma_i^x, \sigma_j^y\right)$, where the superscript helps to distinguish the series.

$$g(Z_s) = (f(x_s), f(y_s)) = \eta_{ij} = \left(\sigma_i^x, \sigma_j^y\right). \tag{7}$$

We say that s is $\eta_{ij} - type$ for $Z = (x, y)$ or simply that location s is $\eta_{ij} - type$, if and only if s is $\sigma_i^x - type$ for x and $\sigma_j^y - type$ for y.

A natural way to symbolize a series, namely, x, is to embed it in an m − dimensional space as follows:

$$x_m(s) = \left(x_s, x_{s_1}, \ldots, x_{s_{m-1}}\right) \quad for \quad s \in S, \tag{8}$$

where $m \geq 2$ is the embedding dimension and $\{s_j; j = 1, 2, \ldots, m - 1\}$ are the $(m - 1)$ nearest neighbours of s (the user should define the notion of neighborhood). We use the term m − surrounding in reference to $x_m(s)$.

We are now ready to symbolize the series, for which we need a specific symbolization map. According to our experience, the following procedure is simple and efficient for treating with spatial data (Matilla-García and Ruiz Marín 2008, 2009; López et al. 2010; Ruiz et al. 2010 for more patterns).

1. Define the indicator function:

$$\tau_s = \begin{cases} 1 & \text{if} \quad x_s \geq M_e^x \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

   where M_e^x is the median of series x.
2. Define a second indicator

$$\iota_{ss_i} = \begin{cases} 0 & \text{if} \quad \tau_s \neq \tau_{s_i} \\ 1 & \text{otherwise} \end{cases}, \tag{10}$$

   ι_{ss_i} relates each location, s, with every other location of its m − surrounding s_i, $i = 1, 2, \ldots, m - 1$.
3. The symbolization map for $\{x_s\}_{s\in S}$ is

$$f(x_s) = \sum_{i=1}^{m-1} \iota_{ss_i}. \tag{11}$$

Note that, in this case, the cardinality of the set of symbols is m (that is, $m = n$ and $\Gamma_m = \{0, 1, \ldots, m - 1\}$).

The final result is simply a natural number (the symbol) assigned to each location s that, according to the previous symbolization map, measures the degree of coincidence of variable x in location s with its neighbourhood. The same can be done for y and, finally, also for the bivariate process Z.

3.2 Entropy: Definitions and concepts

The symbols entail a (rough) description of the spatial distribution of the associated process, and entropy is a measure of the information conveyed by a distribution function. It seems natural to relate the two concepts in what we call symbolic entropy. We begin by introducing some basic definitions.

Let us assume that a is a discrete random variable that takes on values $\{a_1, a_2, \ldots, a_n\}$ with probabilities $p(a_i)$ for $i = 1, 2, \ldots, n$. Shannon entropy for this variable, $h(a)$, is defined as:

$$h(a) = -\sum_{i=1}^{n} p(a_i)\ln(p(a_i)).$$

The extension to the bivariate case (a, b), where b is another discrete variable with a finite number of categories, is simple: $h(a,b) = -\sum_a \sum_b p(a,b)\ln(p(a,b))$. In this case, we need the joint probability distribution $p(a, b)$. Similarly, we can obtain the conditional entropy of variable a with respect to b: $h(a|b) = -\sum_a \sum_b p(a,b)\ln(p(a|b))$. The conditional entropy measures the entropy of a that remains when b has been observed, and it is very important in our procedure because it encapsulates the notion of additional information upon which rests Granger's approach.

It is true that the probabilities of the symbols are unknown, but they can be estimated using the corresponding frequencies. After symbolizing the series $\{x_s\}_{s\in S}$ and $\{y_s\}_{s\in S}$ for a given embedding dimension, $m \geq 2$, we calculate the absolute and relative frequencies of the set of symbols σ_i^x:

$$n_{\sigma_i^x} = \#\{s \in S \mid s \quad \text{is} \quad \sigma_i^x - type \quad \text{for} \quad x\}, \tag{12}$$

$$p(\sigma_i^x) \equiv p_{\sigma_i^x} = \frac{\#\{s \in S \mid s \quad is \quad \sigma_i^x - type \quad for \quad x\}}{|S|} = \frac{n_{\sigma_i^x}}{|S|}, \tag{13}$$

where $|S|$ denotes the cardinality of set S, in general $|S| = N$. Similarly for the symbols corresponding to y, σ_j^y, and for those of the bivariate case, $Z_s = \{x_s, y_s\}_{s\in S}$, $\eta_{ij} \in \Omega_n$:

$$p(\eta_{ij}) \equiv p_{\eta_{ij}} = \frac{\#\{s \in S \mid s \quad is \quad \eta_{ij} - type\}}{|S|} = \frac{n_{\eta_{ij}}}{|S|}. \tag{14}$$

The symbolic entropy for the *two–dimensional* spatial series $\{Z_s\}_{s\in S}$ is the Shannon entropy:

$$h_Z(m) = -\sum_{\eta \in \Omega_m^2} p(\eta)\ln(p(\eta)). \tag{15}$$

Similarly, we estimate the marginal symbolic entropies as:

$$h_\delta(m) = -\sum_{\sigma^\delta \in \Gamma_m} p(\sigma^\delta)\ln(p(\sigma^\delta)), \delta = x, y. \tag{16}$$

The symbolic entropy of y, conditioned to symbol σ^x in x is:

$$h_{y|\sigma^x}(m) = -\sum_{\sigma^y \in \Gamma_m} p(\sigma^y \mid \sigma^x) \ln\left(p(\sigma^y \mid \sigma^x)\right), \tag{17}$$

from which we estimate the conditional symbolic entropy of y given x:

$$h_{y|x}(m) = -\sum_{\sigma^x \in \Gamma_m} \sum_{\sigma^y \in \Gamma_m} p(\sigma^x, \sigma^y) \ln\left(p(\sigma^y \mid \sigma^x)\right). \tag{18}$$

After a few manipulations, and using the well-known result that $p(\sigma^x, \sigma^y) = p(\sigma^x)p(\sigma^y|\sigma^x)$, we obtain a more compact expression for the conditional symbolic entropy:

$$h_{y|x}(m) = \sum_{\sigma^x \in \Gamma_m} p(\sigma^x) h_{y|\sigma^x}(m). \tag{19}$$

All the elements that appear in (19) have been estimated previously, in (13) or in (17), so this magnitude can also be estimated. As indicated, the conditional symbolic entropy of y given x measures the uncertainty that remains about y, once the distribution of x, has been fixed. In causal terms, a decrease in this quantity means a certain impact of x on y.

4 Spatial causality in information

The test to be present in this Section will help us to determine if a variable, x, causes another variable, y, according to the spatial patterns observed for the two variables. It is worth remembering that this test is the final stage of a longer process. As discussed in Herrera (2011), a number of issues have to be considered:

(1) The role of the space. Two series are analysed in a spatial support and it is crucial making sure that space is a factor that we need to consider and how it should be treated. This amounts to discussing which is the most adequate spatial weight matrix and testing the assumption of spatial independence of both series. If the series are spatially independent, in a univariate sense, a traditional causality analysis would be advisable (i.e., as in Heckman 1999).
(2) The relation between the variables. Assuming that the space is relevant and that we have selected the corresponding weighting matrix, we need to test for the existence of spatial dependence between the two variables (Herrera et al. 2013, for a recent test in this direction). This is a necessary condition for causality.
(3) Assuming that (1) and (2) are satisfied, causality in information implies that there is a one-way information flow between the two variables. Part of the literature on causality (Charemza and Deadman 1997, for example) discards the notion of simultaneous causality (x and y cause each other simultaneously) as a case of effective causation between variables. We follow this approach. By 'one-way information flow' we mean that entropy measures detect causality in one direction but not in the reverse direction.

Below we focus on the third point; the other two are beyond the scope of the current paper.

Let $\{x_s\}_{s \in S}$ and $\{y_s\}_{s \in S}$ be two spatial processes and let $\mathcal{W}(x, y)$ be the set of spatial-dependence structures (that is, the set of spatial weighting matrices relevant for each variable) between x and y.

We use

$$\mathcal{X}_{\mathcal{W}} = \{W_i x \mid W_i \in \mathcal{W}(x, y)\}, \tag{20}$$

$$\mathcal{Y}_{\mathcal{W}} = \{W_i y \mid W_i \in \mathcal{W}(x, y)\}, \tag{21}$$

to denote the sets of spatial lags of x and y given by all the weighting matrices in $\mathcal{W}(x, y)$.

Definition. *We say that* $\{x_s\}_{s\in S}$ *does not cause* $\{y_s\}_{s\in S}$ *under the spatial structures* $\mathcal{X}_{\mathcal{W}}$ *and* $\mathcal{Y}_{\mathcal{W}}$ *if*

$$h_{y|\mathcal{Y}_{\mathcal{W}}}(m) = h_{y|\mathcal{Y}_{\mathcal{W}},\mathcal{X}_{\mathcal{W}}}(m). \tag{22}$$

Then, we propose an unilateral non-parametric test for the following null hypothesis

$$H_0 : \{x_s\}_{s\in S} \text{ does not cause } \{y_s\}_{s\in S} \text{ under the spatial structures } \mathcal{X}_{\mathcal{W}} \text{ and } \mathcal{Y}_{\mathcal{W}}, \tag{23}$$

with the following statistic:

$$\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}) = \hat{h}_{y|\mathcal{Y}_{\mathcal{W}}}(m) - \hat{h}_{y|\mathcal{Y}_{\mathcal{W}},\mathcal{X}_{\mathcal{W}}}(m). \tag{24}$$

That is, if $\mathcal{X}_{\mathcal{W}}$ does not contain extra information about y then $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}) = 0$, otherwise, $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}}) > 0$. As usual, the hat, '^', means estimated. The alternative is that the null hypothesis of (23) is not true.

Note that, as indicated in subsection 3.2, $h_{y|\mathcal{Y}_{\mathcal{W}}}(m)$ measures the uncertainty of the distribution of symbols of y, conditional to the symbols of its spatial lag, $\mathcal{Y}_{\mathcal{W}}$. Moreover $h_{y|\mathcal{Y}_{\mathcal{W}},\mathcal{X}_{\mathcal{W}}}(m)$ measures the uncertainty of the distribution of symbols of y, conditional to the symbols of the spatial lags of y, $\mathcal{Y}_{\mathcal{W}}$, and of x, $\mathcal{X}_{\mathcal{W}}$. If the second variable, x, indeed causes the first one then there should be a significant decrease in the entropy, and the statistic of (24) will take on high positive values. On the contrary, if there is only correlation, but not causation, the difference between both entropies will be small.

In order to remain in a model-free framework, we prefer to determine the significance of the test using bootstrap methods. Following the guidelines of the non-overlapping time block bootstrap of Carlstein (1986), Herrera et al. (2013) apply a similar approach in spatial context. Under the name of spatial block bootstrap, SBB, the authors break down the dependence between the series but preserving most of the spatial structure in each series. The SBB procedure, with a number B of bootstraps, consists of the following steps:

1. Compute the value of the statistic $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ using the original data, $\{x_s\}_{s\in S}$ and $\{y_s\}_{s\in S}$.
2. Divide each spatial series into $b = N/l$ contiguous observational blocks of l units. Remember that N is the sample size. By contiguous observational blocks we mean that the observations of each block are contiguous according to the W matrix. The blocks cannot overlap and must cover the entire space.[1]
3. Generate two new samples of length N by resampling, with replacement, the b blocks of x and y. Let us call $\{x_s(i)\}_{s\in S}$ and $\{y_s(i)\}_{s\in S}$ the two bootstrapped series, where i is the number of the bootstrap sample.

[1] This condition means that each location belongs to one and only one block and the blocks are designed to contain and exhaust all locations.

4. Estimate the bootstrapped realization of the statistic $\hat{\delta}^{(i)}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ using the bootstrapped series $\{x_s(i)\}_{s\in S}$ and $\{y_s(i)\}_{s\in S}$.
5. Repeat $B-1$ times steps 3 and 4 to obtain B bootstrapped realizations of the statistic $\left\{\hat{\delta}^{(i)}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right\}_{i=1}^{B}$.
6. Compute the estimated bootstrap $p-$ value:

$$p_{boots}-value\left(\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right)=\frac{1}{B}\sum_{i=1}^{B}\tau\left(\hat{\delta}^{i}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})>\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right), \tag{25}$$

where $\tau(\cdot)$ is an indicator function that assigns 1 if inequality is true and 0 otherwise.
7. Reject the null hypothesis that $\{x_s\}_{s\in S}$ does not cause $\{y_s\}_{s\in S}$ under the spatial structure $\mathcal{W}(x, y)$ if

$$p_{boots}-value\left(\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})\right)<\alpha, \tag{26}$$

for a nominal size α.

5 Monte Carlo simulations

The objective of this section is to study the behaviour of the test in (24) for finite samples. We examine the empirical size in two cases: (*i*) when x and y are *iid* and (*ii*) if they are spatially dependent. We also want to examine the power of the test in presence of linear and non-linear spatial causality.

5.1 Experimental design

Each experiment begins by obtaining a random map in a bivariate system of co-ordinates. Then a normalized W matrix is built following the $m-1$ nearest neighbours criterion.

The following global parameters are involved in the data generating process (DGP):

$$N\in\{320, 560, 800\}, m\in\{4, 5, 6\}, \tag{27}$$

where N is the sample size and m is the embedding dimension. We simulate linear and non-linear relations between the two variables x and y:

$$y=(I-\rho W)^{-1}(\beta x+\theta Wx+\varepsilon), \tag{28}$$

$$y=\exp\left[(I-\rho W)^{-1}(\beta x+\theta Wx+\varepsilon)\right], \tag{29}$$

where $x\sim\mathcal{N}(0,1), \varepsilon\sim\mathcal{N}(0,1)$, $\text{cov}(x, \varepsilon)=0$ and $\rho\in\{0.0; 0.3; 0.5; 0.7\}$.

We want to control for the intensity of the signal between the two variables using the expected $R^2_{y/x}$ coefficient in a hypothetical linear equation:

$$y=\beta x+\theta Wx+\varepsilon. \tag{30}$$

The expected $R^2_{y/x}$ in (30), under the assumptions above, is:

$$R^2_{y/x}=\frac{\beta^2+(\theta^2 m-1)}{\beta^2+(\theta^2 m-1)+1}.$$

We have considered two values for this coefficient:

$$R^2_{y/x} \in \{0.6; 0.8\}. \tag{31}$$

For simplicity, in all cases we use $\beta = 0.5$. The spatial lag parameter of x is obtained simply as: $\theta = \sqrt{\frac{(1-m)(\beta^2(1-R^2)-R^2)}{1-R^2}}$.

The empirical size has been estimated using two independent processes such as:

$$y = \rho_y Wy + \varepsilon_y, \tag{32}$$

$$x = \rho_x Wx + \varepsilon_x, \tag{33}$$

where as before $\varepsilon_\lambda \sim \mathcal{N}(0,1); \lambda = x, y$ and:

$$\rho_y \in \{0.0; 0.4; 0.8\},$$

$$\rho_x \in \{0.0; 0.3; 0.7\}.$$

As a general rule, following Rohatgi (1976), we consider that there should be, on average, five observations for each symbol whose frequency is estimated (this decision affects the sample size and the number of variable used in the experiment).

5.2 *Monte Carlo results*

The size is the percentage of false rejections of the null hypothesis of no causality from x to y. Under the best of circumstances, this empirical size should be closed to the nominal level. If the empirical size is smaller than the nominal level, the test is called conservative, and oversized in the opposite case.

It is important to stress that all the results below have been obtained after a double testing sequence. That is, global empirical size reflects the percentage of cases in which the hypothesis of (23) is incorrectly rejected and, simultaneously, its complement (y does not cause x) is correctly non-rejected. That is, we compute a false positive only if the hypothesis of (23) is unambiguously rejected by detecting one way information flow from x to y. Similarly, global estimated power is the percentage of rejections of the null of (23) with the simultaneous non-rejection of the complementary hypothesis (y does not cause x). Simultaneous rejection of both null hypothesis, as indicated before, does not allow identifying an unambiguous direction of causality. These cases are excluded when computing the estimated power.

Table 1 shows the empirical size of the statistic at a 5 per cent nominal level. As it is evident from this table, the best situation for the $\hat{\delta}(\mathcal{Y}_\mathcal{W}, \mathcal{X}_\mathcal{W})$ test corresponds to *iid* series ($\rho_y = \rho_x = 0$) when the empirical size is close to the nominal level. The test becomes conservative for higher values in any of the two parameters of spatial dependence. In other words, it is more difficult to reject the null hypothesis of non causality when there is spatial structure in the variables.

Table 2 shows the estimated global power obtained for the linear case (28). As expected, the estimated power function increases with the size of the sample size and with the embedding dimension. For high values of the coefficient of spatial autocorrelation of y, such as 0.7, and small sample sizes the estimated power decreases slightly. Note that the estimated power for large samples and ample m – surroundings is 100 per cent.

Table 1. Global empirical size of $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ test at 5 per cent level

ρ_y	ρ_x	$N = 320$	$N = 560$		$N = 800$		
		$m = 4$	$m = 4$	$m = 5$	$m = 4$	$m = 5$	$m = 6$
0.0	0.0	5.5	4.9	5.4	5.5	3.4	4.0
0.0	0.3	3.8	3.7	4.9	3.6	4.2	3.1
0.0	0.7	2.0	1.5	1.1	4.0	2.2	0.3
0.4	0.0	3.7	4.7	3.0	5.2	3.2	2.4
0.4	0.3	2.7	2.9	2.1	3.0	2.1	1.7
0.4	0.7	1.5	2.5	0.4	2.5	1.2	0.2
0.8	0.0	1.0	1.6	0.9	2.9	0.6	0.4
0.8	0.3	0.8	1.8	0.5	1.8	0.5	0.3
0.8	0.7	0.2	0.7	0.1	1.0	0.1	0.1

Notes: Boots: 399. Blocks: 8. Replications: 1,000.

Table 2. Global estimated power of $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ test at 5 per cent level. Equation (28): linear case

N	ρ	$R^2_{y/x} = 0.6$			$R^2_{y/x} = 0.8$		
		$m = 4$	$m = 5$	$m = 6$	$m = 4$	$m = 5$	$m = 6$
320	0.3	92.2	–	–	87.5	–	–
320	0.5	97.0	–	–	97.3	–	–
320	0.7	92.2	–	–	95.8	–	–
560	0.3	87.6	98.3	–	82.0	98.0	–
560	0.5	96.1	99.6	–	94.5	99.3	–
560	0.7	99.3	99.7	–	98.8	99.8	–
800	0.3	68.0	93.3	99.4	54.6	90.1	99.7
800	0.5	87.5	98.6	100.0	83.1	98.6	99.9
800	0.7	96.9	99.8	100.0	95.9	99.7	100.0

Notes: Boots: 399. Blocks: 8. Replications: 1,000. Empty cells correspond to cases where number of observations per symbol is below of 5.

Table 3. Global estimated power of $\hat{\delta}(\mathcal{Y}_{\mathcal{W}}, \mathcal{X}_{\mathcal{W}})$ test at 5 per cent level. Equation (28): non-linear case

N	ρ	$R^2_{y/x} = 0.6$			$R^2_{y/x} = 0.8$		
		$m = 4$	$m = 5$	$m = 6$	$m = 4$	$m = 5$	$m = 6$
320	0.3	91.3	–	–	87.7	–	–
320	0.5	95.5	–	–	94.6	–	–
320	0.7	93.8	–	–	95.4	–	–
560	0.3	82.3	93.0	–	78.8	92.7	–
560	0.5	91.7	98.3	–	87.6	97.4	–
560	0.7	97.2	99.5	–	96.1	99.6	–
800	0.3	69.5	87.9	97.0	57.2	82.4	95.7
800	0.5	83.3	94.8	99.3	78.9	93.7	99.3
800	0.7	94.9	98.3	100.0	91.0	97.9	99.8

Notes: Boots: 399. Blocks: 8. Replications: 1,000. Empty cells correspond to cases where number of obs. per symbol is below of 5.

For the non-linear process, Table 3, the test shows a good performance in almost all cases. As before, the estimated power improves with the sample size and the dimension of the m – surroundings.

Overall, these results are quite satisfactory in spite of the test being conservative. A conservative test reduces the possibility of committing the error of correctly rejecting the null hypothesis but for the wrong reasons. The situation of the $R^2_{y/x}$ is a bit puzzling. Its impact is positive for small samples (higher $R^2_{y/x}$ means, in general, higher power) but a worsening effect, due to the univariate spatial dependence of the series, tend to offset when the sample size increases. We are working in this aspect.

6 Unemployment and migration in US counties

The debate between unemployment and migration has a long tradition in the economic literature. At the risk of simplifying, we can say that the axiom that immigration causes unemployment is part of the neoclassical paradigm (Borts and Stein 1964). Assuming homogeneity in the labour force and perfect competition in the market of goods, workers move to prosperous regions attracted by higher salaries. The inflow increases labour supply in these regions (direct effect). In turn, migrants increase the consumption of local goods, leading companies to hire more workers (indirect effect). It is customary to assume that direct effects prevail over the indirect effects, raising unemployment in the regions of destination. New economic geography (Krugman 1991) assumes imperfect competition on the goods market and rigid labour markets but the conclusion is the same: migration causes unemployment.

Other strands of literature conclude that unemployment is the cause of migration. Pissarides and Wadsworth (1989) argue that people move from places where they are not fully employed to places offering greater possibilities of being employed. In fact, it is well documented that unemployment in the place of origin increases the willingness to migrate (Antolin and Bover 1997).

In sum, this controversy constitutes an interesting piece of research because:

- There is no consensus about the relation between unemployment and migration; let us add that there exist abundant evidence supporting both hypothesis.
- The spatial dimension, although not in a very prominent role, has been frequently considered in this debate.
- Typical data consist of national, or regional, time series data; rarely a pure cross-section has been used.
- Granger notion is the preferred option in causality studies in this field.

This section analyses the relation between unemployment and net migration in 3,108 US counties for the period 2003–2008. The purpose is testing for causality between the two variables and, if so, detecting the direction of causation using the methodology introduced previously.

We use annual data. The unemployment rate is equal to the number of unemployed divided by the labour force.[2] Net migration is the difference between in-migration to an area and out-migration from the same area, on a July-to-July comparison, as a proportion of an area's population at the midpoint of the time period and expressed per 1,000 population.[3]

Table 4 presents a summary for the 3,108 US counties during the six years period. The unemployment rate fell for the whole period 2003–2007. The average rate for the counties was

[2] Unemployment rate was estimated using labour force data by county, annual averages (Bureau of Labor Statistics, Local Area Unemployment Statistics programme).

[3] Includes domestic and international migration (United States Census Bureau, Demographic Components of County Population Change).

Table 4. Unemployment rate and net migration by US county

Year	Unemployment Rate (%)		Net Migration (‰)	
	Average	Stand. Dev.	Average	Stand. Dev.
2003	5.97	1.93	0.85	13.83
2004	5.63	1.77	1.63	15.34
2005	5.38	1.78	1.28	15.39
2006	4.89	1.67	2.26	18.44
2007	4.84	1.69	1.15	14.77
2008	5.76	2.07	–0.14	14.42

Source: US Census, Bureau of Labor Statistics.

5.97 per cent in 2003, falling to 4.84 per cent in 2007. Net migration grew in the first four periods, from an average of 0.85 per cent in 2003 to 2.26 per cent in 2006. Years 2007 and 2008 show a decrease in the aggregated value reaching a negative mean to –0.14 per cent, that is outmigration exceed immigration on average. The dispersion differs among variables and between periods. In the case of unemployment, the standard deviation fell from 2003 to 2006 and then increased over 2008 reaching a value of 2.07. Dispersion in the net migratory rate grew throughout the first four periods but decreased for the last two years.

Net migration is measured from July to July but unemployment is measured on a calendar year. In order to reduce this mismatch, we use averaged data for 3-years periods: 2003–2005, 2004–2006, 2005–2007, 2006–2008, plus the average corresponding to the overall period 2003–2008.

Figure 3 shows the spatial distribution of unemployment and net migration (averaging all years). There are clusters of high unemployment in West Coastal counties and in some North-eastern regions in the area of Michigan, Wisconsin and Maine. Furthermore, we can detect clusters of low unemployment rate in North Central regions and in some states of East Coast (Virginia, Maryland, Columbia, Vermont and New Hampshire). With regards to migration flows, the Western and Eastern states are recipients whereas the central states are generators of migrants.

As said, there are 3,108 counties in the sample. Applying the rule of $m^3 \cdot 5 \cong N$, we choose an embedding dimension m equal to 9, which means that the symbol corresponding to each county has been obtained using the eight nearest neighbours.

Before testing causality, we did some preliminary checks in order to assure that the weighting matrix corresponding to the eight nearest neighbours is an adequate selection. Moreover, there is strong spatial dependence in these series according to Moran's I. Finally, and conditional to the spatial structure identified in the weighting matrix, unemployment and migration are statistically dependent over space, as shown by the Ψ_2 test (Herrera et al. 2013). Main results appear in Table 5.

Table 6 shows the results of the testing strategy introduced in Section 4. We detect spatial causality from net migration to unemployment. Indeed, there is a clear signal of causality in our county dataset, in the sense that the information flow is unidirectional from migration to unemployment.

Summarizing, space is relevant to interpret the relation between net migration and unemployment in the US counties; the two variables are not spatially independent and there exists causality from migration to unemployment. The reverse causal mechanism is not supported by the data.

Unemployment must be on the left hand side and migration on the right hand side of the equation. Then we can proceed with the labour of building the model in a gets or stge strategy.

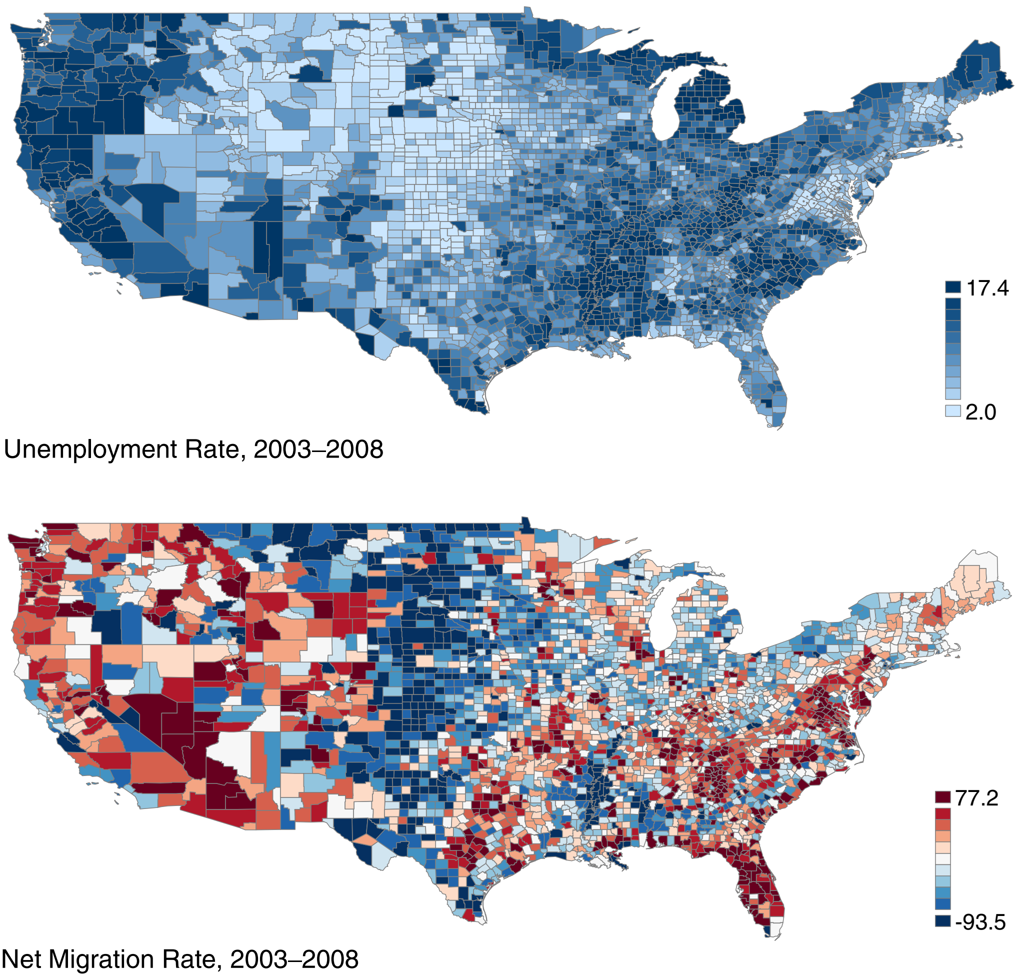

Fig. 3. Spatial distribution of unemployment and net migration by county. Average 2003–2008

Table 5. Spatial dependence tests

Year	Unemploment	Net Migration	Net Migration – Unemployment
	Moran's I	Moran's I	p – value (Ψ_2)
2003–2005	0.55***	0.40***	0.030**
2004–2006	0.56***	0.39***	0.047**
2005–2007	0.58***	0.38***	0.040**
2006–2008	0.61***	0.36***	0.027**
2003–2008	0.58***	0.42***	0.045**

Notes: **: significant at 5%, ***: significant at 1%. Permuts: 499, *W*: 8 nearest neighs.

Let us assume that we cannot reject linearity and that, finally, a spatial Durbin model has been selected, such as:

$$y_i = \rho \mathbf{w}_i' y + \beta_1 x_i + \beta_2 \mathbf{w}_i' x + u_i, \tag{34}$$

Table 6. Results of spatial causality test

H_0	Unemploment $\nRightarrow$ Migration	Migration $\nRightarrow$ Unemploment	Conclusion
Periods	p – value	p – value	
2003–2005	0.100	0.045	$Migr. \Rightarrow Unemp.$
2004–2006	0.882	0.048	$Migr. \Rightarrow Unemp.$
2005–2007	0.827	0.049	$Migr. \Rightarrow Unemp.$
2006–2008	0.812	0.047	$Migr. \Rightarrow Unemp.$
2003–2008	0.652	0.050	$Migr. \Rightarrow Unemp.$

Notes: '$\nRightarrow$' means does not cause and '$\Rightarrow$' means causes. $m = 9$, Boots: 399, Blocks: 42 (74 observations by block).

i indexes counties, y is unemployment rate, and x is net migration. Also, u_i an error term, ρ is a parameter that captures the diffusion effect, β_1 is a parameter that captures the *causal effect* of net migration in the location and β_2 is the causal effect of the net migration of the nearest neighbours. After estimating and calibrating the equation, the subsequent study of multipliers should be done to complete the analysis. The difference is that, now, the direct and indirect effects will have a strong causal meaning.

To open the path to other applications, we develop a Matlab routine for this estimation procedure, which can be downloaded for free from the web pages Spatial Causality (https://sites.google.com/site/spatialcausality/codes) and Runmycode (http://www.runmycode.org/companion/view/708).

7 Conclusion and discussion

Pagan (1989, p. 89) admitted his disappointment with the notion of causality in the field of econometrics, writing that 'there was a lot of high powered analysis of this topic, but I came away from a reading of it with the feeling that it was one of the most unfortunate turnings for econometricians in the last two decades, and it has probably generated more nonsense results than anything else during that time'. We partially share his disappointment. However we do think that causality is one of the major concepts developed in modern economics and, because of that, we need powerful techniques to deal with it.

Following the Granger-Wiener tradition, we identify causality with the principle of incremental informative content. Granger's approach is greatly dependent of the temporal precedence principle: there must be a time lag between cause and effect. This is intuitive but it is not applicable to pure cross-sections, where the notion of additional informative content works better. We have tried to extend this reasoning by developing adequate measures of information content, with the aim of solving the causality analysis. Intuitively, our definition establishes that the cause variable should provide additional, and unique, information about the effect variable.

The δ test of spatial causality developed in this paper compares two distribution functions; in fact it looks for significant differences in their informative content by using measures of conditional entropy. One of the distributions uses all the information available in the space concerning the effect variable, whereas the other adds the spatial information of the variable assumed to be the cause. The test is intuitive and does not need of any hypothesis about functional form, distribution function, or other aspects of the specification. It is a fully non-parametric causality test.

This proposal sharply contrasts to the usual practice in applied literature, where dominates an approach based on the Wold decomposition theorem and the VAR representation of stationary time series. The extension of VAR methodology to a spatial framework has been subject to intense investigation in recent years with significant results (i.e., Beenstock and Felsenstein 2007; Mutl 2009; Di Giacinto 2010, to mention just a few). Indeed, it offers some advantages (simplicity, easy to generalize, etc.) but also has weaknesses among which we can mention: (*i*) spatial data are usually more complicated than time series data, at least because of the multidirectionality of the relations and to the uncertainty in relation to the so-called weighting matrix; and (*ii*) the notion of prediction in a spatial setting is difficult to apply (in spite of some recent interesting work, Kelejian and Prucha 2007). Specifically, research made by the authors in this direction did not produce acceptable results with respect to spatial causality. In any case, this is a field that merits more investigation in the future.

One of the problems of our own proposal is that we need of symbolization maps to evaluate and compare the entropies of the distributions. We distinguish between standard and non-standard symbolization functions (the first assigns equal probability to each symbol, whereas the second allows for symbols with different probabilities). Standard symbolization maps were used previously to capture spatial dependence in quantitative and/or qualitative series (see López et al. 2010; Ruiz et al. 2010). The symbolization function developed in our paper is non-standard, partly, because this type of function assures a more efficient use of the information to estimate the probability of the symbols.

It should be acknowledged that the discussion in relation to the symbolization function (specification, properties, comparison, selection, etc.) is an area still under development that needs further research. The function used in this paper produced reasonable results, but we cannot exclude the existence of better optimal symbolization functions.

Another difficulty related to our approach is the generalization to the case of more than two variables. From a theoretical point of view, the extension to a multivariate context is not a problem: it is enough with extending the set of conditioning factors that appear in the entropy of the right hand side of (22), (i.e., the trivariate case where z and x are the variables assumed cause of y, $h_{y|\mathcal{Y}_W,\mathcal{X}_W,\mathcal{Z}_W}$, where the notation is obvious). The real problem in this direction is the inflation of degrees of freedom: the more potential causing variables we introduce in the analysis, the more symbols we need to add to the procedure, which increases the need of observations to estimate correctly their probability. This point is clearly related with the problem of producing efficient symbolization functions.

Finally, let us remind that the δ test shows a good behaviour in samples of medium to large sample size. Moreover, it is robust to the functional form of the relation and we do not need the assumption of normality. This procedure to detect causality has been applied to the debate about unemployment versus migration. There is an abundant literature supporting the assumptions that migration causes unemployment and also that unemployment causes migration. Using data on US counties for the 2003–2008 period, we have found clear evidence supporting the first hypothesis: migration causes unemployment.

It is clear that the topic of causality in spatial econometric models requires more attention. More work is needed in this field, both of an applied and methodological type. Our impression is that the topic is of sufficient importance to justify this effort.

References

Angrist J, Pischke J (2009) *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, Princeton, NJ

Anselin L (1988) *Spatial econometrics: Methods and models*, Volume 4. Kluwer Academic Publishers, Dordrecht

Antolin P, Bover O (1997) Regional migration in Spain: The effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections. *Oxford Bulletin of Economics and Statistics* 59: 215–235

Beenstock M, Felsenstein D (2007) Spatial vector autoregressions. *Spatial Economic Analysis* 2: 167–196

Borts GH, Stein JL (1964) *Economic growth in a free market*. Columbia University Press, New York
Brockwell PJ, Davis RA (2002) *Introduction to time series and forecasting*, Volume 1. Springer-Verlag, New York
Carlstein E (1986) The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *The Annals of Statistics* 14: 1171–1179
Cartwright N (1995) Probabilities and experiments. *Journal of Econometrics* 67: 47–59
Charemza W, Deadman DF (1997) *New directions in econometric practice: General to specific modelling, cointegration and vector autoregression*. Edward Elgar, Cheltenham
Davidson J (2000) *Econometric theory*. Blackwell Publishers, Oxford
Di Giacinto V (2010) On vector autoregressive modeling in space and time. *Journal of Geographical Systems* 12: 125–154
Gibbons SH, Overman G (2012) Mostly pointless spatial econometrics? *Journal of Regional Science* 52: 172–191
Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–438
Hao B-L, Zheng W-M (1998) *Applied symbolic dynamics and chaos*, Volume 7. World Scientific, Singapore
Heckman JL (1999) Causal parameters and policy analysis in economics: A twentieth century retrospective. Technical report, National Bureau of Economic Research
Herrera M (2011) Causality: Contributions to spatial econometrics. Ph. D. thesis, University of Zaragoza
Herrera M, Ruiz M, Mur J (2013) Detecting dependence between spatial processes. *Spatial Economic Analysis* 8: 469–497
Hlavácková-Schindler K (2009) Causality in time series: Its detection and quantification by means of information theory. In: Emmert-Streib F, Dehmer M (eds) *Information theory and statistical learning*. Springer, New York
Hoover KD (2004) Lost causes. *Journal of the History of Economic Thought* 26: 149–164
Isard W (1971) *Métodos de análisis regional: Una introducción a la ciencia regional*, Volume 1. Editorial Ariel, Barcelona
Isserman AM, Merrifield J (1982) The use of control groups in evaluating regional economic policy. *Regional Science and Urban Economics* 12: 43–58
Kelejian HH, Prucha IR (2007) The relative efficiencies of various predictors in spatial econometric models containing spatial lags. *Regional Science and Urban Economics* 3: 363–374
Krugman P (1991) Increasing returns and economic geography. *The Journal of Political Economy* 99: 483–499
LeSage J, Pace R (2009) *Introduction to spatial econometrics*. CRC press, Boca Raton FL
López F, Matilla-García M, Mur J, Marín MR (2010) A non-parametric spatial independence test using symbolic entropy. *Regional Science and Urban Economics* 40: 106–115
Matilla-García M, Ruiz Marín M (2008) A non-parametric independence test using permutation entropy. *Journal of Econometrics* 144: 139–155
Matilla-García M, Ruiz Marín M (2009) Detection of non-linear structure in time series. *Economics Letters* 105: 1–6
Mutl J (2009) Panel VAR models with spatial dependence. Technical report, Reihe Ökonomie/Economics Series, Institut für Höhere Studien (IHS) 237
Pagan A (1989) 20 years after: Econometrics 1966–1986. In: Cornet B, Tulkens H (eds) *Contribution to operations research and econometrics*. The MIT Press, Cambridge
Partridge MD, Boarnet M, Brakman S, Ottaviano G (2012) Introduction: Whither spatial econometrics? *Journal of Regional Science* 5: 167–171
Pearl J (2009) *Causality: Models, reasoning and inference* (2nd ed.), Volume 29. Cambridge University Press, Cambridge
Pissarides CA, Wadsworth J (1989) Unemployment and the inter-regional mobility of labour. *The Economic Journal* 99: 739–755
Rohatgi V (1976) *An introduction to probability theory and mathematical statistics*. John Wiley, Chichester
Ruiz M, López F, Páez A (2010) Testing for spatial association of qualitative data using symbolic dynamics. *Journal of Geographical Systems* 12: 281–309
Sims CA (1972) Money, income, and causality. *The American Economic Review* 62: 540–552
Wiener N (1956) The theory of prediction. In: Beckenbach EF (ed) *Modern mathematics for the engineer*. McGraw-Hill, New York
Zellner A (1988) Causality and causal laws in economics. *Journal of econometrics* 39: 7–21

doi:10.1111/pirs.12144

Resumen. Este artículo se centra en el problema de probar la existencia y la dirección de la causalidad en un grupo de variables, mediante un marco espacial. Nuestro objetivo es producir criterios útiles para debatir la causalidad en función de bases racionales y objetivas. Se introduce un nuevo test no paramétrico, basado en la entropía simbólica, que es robusto en cuanto a la forma funcional de la relación. La prueba muestra un buen comportamiento en muestras de tamaño medio a grande. La propuesta se ilustra con un ejemplo en torno al caso de la migración frente al desempleo, a partir de datos de 3108 condados de los EE.UU. para el período 2003–2008.

要約: 本論文は、空間的フレームワークを用いて、変数グループの存在と因果関係の方向を検定する問題に焦点を当てる。我々の目的は、合理的および客観的根拠により、因果関係を検討する有効な基準を構築することである。我々は、シンボリック・エントロピーに基づいた新しいノンパラメトリック検定を導入する。この検定法は、関係の関数形式に対して頑健であり、中規模ならびに大規模の標本サイズに対して実効的である。2003年から2008年の米国3,108郡のデータによる移住対失業率の事例に適用して、プロポーザルを解説する。