Symbolic Dynamic to Test Basic Hypothesis in Panel Data

Abstract

The purpose of this paper is to show the capacity of a new approach based on the symbolic permutation entropy to deal with different types of data: time series, spatial series or panel data. We focus the attention on the panel data case. For this type of variables, we present a unified non-parametric framework in which we solve various inference problems related to the stochastic structure of the data as, for example, the analysis of the assumption of independence or the detection of structural breaks.

1 Introduction

The interest of combining cross-sectional with time series in panel data sets has been always present in the econometric literature (Nerlove, 2002, for an historic perspective). The strong interest of the topic explains the vast literature devoted to panel data models (Arellano, 2003, or Hsiao, 2003, for two recent textbooks).

In this context, we propose a new, nonparametric approach that may of great help. Initially we focus the discussion on the hypothesis of absence of any spatiotemporal structure in the panel data set, that is, on the null of independence. However, this is only one part of the story because our proposal is, indeed, much more ambitious. With small adjustments, we can also treat with different problems of interest in order to decide how to model a spatial panel data set. For example, the assumption of parameter stability or, more generally, the maintenance of the same stochastic structure among the different cross-sections (or individuals) can be easily solved under our setting. The method that we present is based on symbolic entropy (see Joe, 1989a and b, Hong and White, 2005, for an overview of entropy based tests), a very flexible and powerful non-parametric technique that has been successfully applied in different scenarios (Maasoumi, 1993, Ullah, 1993). To the best of our knowledge, this is the first time that this approach is used in a context of spatiotemporal data with the purpose of testing.

Without loss of generality, we begin the discussion by assuming a family of general null hypotheses that affect to the data generating process of the variable (DGP in what follows). After appropriately symbolizing the data, we translate the problem of testing the hypothesis into the observed distribution of the symbols for the case at hand. The likelihood ratio statistic that is obtained from this distribution plays a crucial role in our procedure. As said, flexibility is one of the main features of this procedure. Indeed, the likelihood ratio may be converted into a test of serial independence for a time series (as in the case of the G test of Matilla and Ruiz, 2008), in a test of cross-sectional independence for a spatial series (as in the case of the SG test of López et al, 2009) or into a test of spatiotemporal independence in a panel data variable as done below in the present paper and which we called STG test. As its predecessors, the STG is a nonparametric test, not very demanding in terms of a priori assumptions. The symbolization that we propose for testing the null hypothesis of spatiotemporal independence assures that it is consistent and invariant to any monotonous transformation of the data and the asymptotic distribution function of the test is standard. To these features we would like to mention that it is easy to obtain and that appears to be well-behaved in term of size and power.

One of the most important features of our approach is that, modifying the symbolization procedure, we can handle with different kind of hypothesis related to the DGP such as, for example, the existence of a structural break in the mechanisms of cross-sectional dependence of the panel data. In fact, the set of symbols used to analyze the cross-sectional structure constitute the starting point for a new test of structural change (called SC). At it is shown in the paper, it is straightforward to use the SC statistic as a test of uniformity between two or more observed spatial distribution (in short, to compare maps of the same variable taken from different regions of the space).

In general, the approach that we present is not conditioned by noisy a priori assumptions. In example, we do not need the restriction of linearity: the STG test is strictly a test of independence; moreover, normality is not a relevant feature for our proposal. Finally and in relation to the spatial dimension of the data, our approach does not require the specification of a weighting matrix which, as indicated by Pinkse (2004), is a not desirable feature.

In the second section we introduce the notation and some basic ideas. The third section presents the general framework in which the likelihood ratio statistic is obtained. Section fourth discusses several symbolization procedures which are adequate for different null hypotheses. The fifth section focuses on obtaining the two tests for independence and structural change in a panel data set. We include also some Monte Carlo evidence. The last section contains some brief conclusions.

2 Preliminaries

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

Let $\{X_{ts}\}_{t\in I, s\in S}$ be a real-valued space-time process, where S is a set of coordinates and I is the time index.

Let $I' \subseteq I$ and $S' \subseteq S$. Let $\Gamma = \{\eta_1, \eta_2, \ldots, \eta_n\}$ be a set of $n > 1$ symbols. Now assume that there exist a map

$$f : \{X_{ts}\}_{t\in I', s\in S'} \to \Gamma.$$

We will say that $(t, s) \in I' \times S'$ is of η_i-type if and only if $f(X_{ts}) = \eta_i, i = 1, 2, ..., n$. We will call the map f a *symbolizing map*. Moreover, if the symbolizing map f is such that under the null of independence all the symbols have the same probability to occur, we will say that f is a *standard symbolizing map*. Otherwise we will say that f is a *non-standard symbolizing map*.

Denote by

$$n_\eta = \sharp\{(t, s) \in I' \times S' |\ f(X_{t,s}) = \eta\},$$

that is, the cardinality of the subset of $I' \times S'$ formed by all the elements of η-type.

Let $|I'| = T$, $|S'| = R$ and $|I' \times S'| = RT$. Then, under these definitions, let us denote by $p(\eta)$ the probability of a symbol. Note that one can easily compute the relative frequency of a symbol $\eta \in \Gamma$ by:

$$p(\eta) := p_\eta = \frac{n_\eta}{RT}. \tag{1}$$

Now under this setting we can define the *symbolic entropy* of a space-time process $\{X_{ts}\}_{t\in I, s\in S}$. This entropy is defined as the Shanon's entropy of the n distinct symbols as follows:

$$h(\Gamma) = -\sum_{\eta\in -} p_\eta \ln(p_\eta). \tag{2}$$

Fix a time period t. Define

$$n_\eta(t) = \sharp\{s \in S' |\ (t, s) \text{ is of } \eta - \text{type}\} \tag{3}$$

and

$$p_\eta(t) = \frac{n_\eta(t)}{RT}. \tag{4}$$

therefore we can restate n_η, the total frequency of a symbol $\eta \in \Gamma$, as:

$$n_\eta = \sharp\{(t, s) \in I' \times S' |\ (t, s) \text{ is of } \eta - \text{type}\} = \sum_{t\in I'} n_\eta(t) \tag{5}$$

and its probability by:

$$p_\eta = \frac{n_\eta}{RT} = \sum_{t \in I'} p_\eta(t) \tag{6}$$

Then, for a fix t, we define the symbolic t-entropy as

$$h_t(\Gamma) = \sum_{i=1}^{n} p_{\eta_i}(t) \ln(p_{\eta_i}(t)). \tag{7}$$

Similarly we define the total entropy of a symbol η as:

$$h(\eta) = \sum_{t=1}^{T} p_\eta(t) \ln(p_\eta(t)). \tag{8}$$

3 The likelihood ratio statistic

In this section we are going to stabilish a framework that enables to construct statistic tests for a family of null hypotheses. To this end we interpret any potential test in terms of symbols' distribution and then the well-known likelihood ratio statistic is used. In other words, any hypothisis test is translated into test on thee symbols'.

Let f be a symbolization map. Now for a symbol $\eta \in \Gamma$ we define the random variable $Z_{\eta ts}$ as follows:

$$Z_{\eta ts} = \begin{cases} 1 & \text{if } f(X_{ts}) = \eta \\ \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

that is, we have that $Z_{\eta ts} = 1$ if and only if (t, s) is of η-type, $Z_{\eta ts} = 0$ otherwise.

Then $Z_{\eta ts}$ is a Bernoulli variable with probability of "success" p_η, where "success" means that (t, s) is of η-type. It is straightforward to see that

$$\sum_{i=1}^{n} p_{\eta_i} = 1 \tag{10}$$

Then we are interested in knowing how many (t, s)'s are of η_i-type for all symbol $\eta \in \Gamma$. In order to answer the question we construct the following variable

$$Y_\eta = \sum_{(t,s) \in I' \times S'} Z_{\eta ts} \tag{11}$$

The variable Y_η can take the values $\{0, 1, 2, \ldots, RT\}$. Then under the assumption of independence among the Z_η's we get that Y_{η_i} is the Binomial random variable

$$Y_\eta \approx B(RT, p_\eta). \tag{12}$$

Then, again under the assumption that the Y_η's are independent, we have that the joint probability density function of the n variables $(Y_{\eta_1}, Y_{\eta_2}, \ldots, Y_{\eta_n})$ is:

$$P(Y_{\eta_1} = a_1, Y_{\eta_2} = a_2, \ldots, Y_{\eta_n} = a_n) = \frac{(a_1 + a_2 + \cdots + a_n)!}{a_1! a_2! \cdot \cdots \cdot a_n!} p_{\eta_1}^{a_1} p_{\eta_2}^{a_2} \cdots p_{\eta_n}^{a_n} \tag{13}$$

where $a_1 + a_2 + \cdots + a_n = RT$. Consequently the joint distribution of the n variables $(Y_{\eta_1}, Y_{\eta_2}, \ldots, Y_{\eta_n})$ is a multinomial distribution.

The likelihood function of the distribution (13) is:

$$L(p_{\eta_1}, p_{\eta_2}, \ldots, p_{\eta_n}) = \frac{RT!}{n_{\eta_1}! n_{\eta_2}! \cdot \cdots \cdot n_{\eta_n}!} p_{\eta_1}^{n_{\eta_1}} p_{\eta_2}^{n_{\eta_2}} \cdots p_{\eta_n}^{n_{\eta_n}} \tag{14}$$

and since $\sum_{i=1}^{n} p_{\eta_i} = 1$ it follows that

$$L(p_{\eta_1}, p_{\eta_2}, \ldots, p_{\eta_n}) = \frac{RT!}{n_{\eta_1}! n_{\eta_2}! \cdot \cdots \cdot n_{\eta_n}!} p_{\eta_1}^{n_{\eta_1}} p_{\eta_2}^{n_{\eta_2}} \cdots (1 - p_{\eta_1} - p_{\eta_2} - \cdots - p_{\eta_{n-1}})^{n_{\eta_n}} \tag{15}$$

Then the logarithm of this likelihood function remains as

$$\begin{aligned} Ln(L(p_{\eta_1}, p_{\eta_2}, \ldots, p_{\eta_n})) &= Ln\left(\frac{RT!}{n_{\eta_1}! n_{\eta_2}! \cdot \cdots \cdot n_{\eta_n}!}\right) + \sum_{i=i}^{n-1} n_{\eta_i} Ln(p_{\eta_i}) \\ &\quad + n_{\eta_n} Ln(1 - p_{\eta_1} - p_{\eta_2} - \cdots - p_{\eta_{n-1}}). \end{aligned} \tag{16}$$

In order to obtain the maximum likelihood estimators of p_{η_i} (i.e., $\widehat{p}_{\eta_i}$) for all $i = 1, 2, \ldots, n$, we solve the following equation

$$\frac{\partial Ln(L(p_{\eta_1}, p_{\eta_2}, \ldots, p_{\eta_n}))}{\partial p_{\eta_i}} = 0 \tag{17}$$

to get that

$$\widehat{p}_{\eta_i} = \frac{n_{\eta_i}}{RT}. \tag{18}$$

Under this setting, if we want to test for a generic null hypothesis, H_0, we will proceed as follows:

1. Fix the null hypothesis H_0 to be tested.
2. Define the set of symbols Γ and the symbolization map f.
3. Compute the distribution of the symbols under H_0, namely $p_\eta^{(0)}$ for all $\eta \in \Gamma$.
4. Finally compute the likelihood ratio statistic.

$$\lambda(Y) = \frac{\frac{RT!}{n_{\eta_1}!n_{\eta_2}!\cdots n_{\eta_n}!} p_{\eta_1}^{(0)n_{\eta_1}} p_{\eta_2}^{(0)n_{\eta_2}} \cdots p_{\eta_n}^{(0)n_{\eta_n}}}{\frac{RT!}{n_{\eta_1}!n_{\eta_2}!\cdots n_{\eta_n}!} \widehat{p}_{\eta_1}^{n_{\eta_1}} \widehat{p}_{\eta_2}^{n_{\eta_2}} \cdots \widehat{p}_{\eta_n}^{n_{\eta_n}}}$$

This general procedure for testing hypotheses based on symbolic analysis will be applied through out the rest of the paper.

Notice that a convenient choice of symbols Γ and symbolization map f in the step 2 according to the null hypothesis to be tested will give an increase in the power of the test. Moreover, $-2Ln(\lambda(Y))$ asymptotically follows a Chi-squared distribution with d degrees of freedom (see for instance Lehmann, 1986). Hence

$$-2Ln(\lambda(Y)) = -2[RTLn(RT) + \sum_{i=1}^{n} n_{\eta_i} Ln\left(\frac{p_{\eta_i}}{n_{\eta_i}}\right)] \sim \chi_d^2 \tag{19}$$

Notice as well that according to Neyman-Pearson's Lemma any test constructed is the most powerful among all level-α tests for this problem.

Now, if the symbolizing map f is *standard*, that is, under the null H_0 all the symbols have the same probability to occur, $p_{\eta_i} = \frac{1}{n}$ for all $i = 1, 2, \ldots, n$, then it follows that

$$\begin{aligned} -2Ln(\lambda(Y)) &= -2RT[Ln(RT) + \sum_{i=1}^{n} \frac{n_{\eta_i}}{RT} Ln\left(\frac{p_{\eta_i}}{n_{\eta_i}}\right)] \\ &= -2RT[Ln(RT) + \sum_{i=1}^{n} \frac{n_{\eta_i}}{RT} (Ln\left(\frac{1}{n}\right) - Ln(n_{\eta_i}))] \\ &= -2RT[Ln(RT) + \sum_{i=1}^{n} \frac{n_{\eta_i}}{RT} (Ln\left(\frac{1}{n}\right) - Ln\left(\frac{n_{\eta_i}}{RT}\right) - Ln(RT))] \end{aligned} \tag{20}$$

Taking into account that $h(\Gamma) = -\sum_{i=1}^{n} p_{\eta_i} \ln(p_{\eta_i}) = -\sum_{i=1}^{n} \frac{n_{\eta_i}}{RT} Ln\left(\frac{n_{\eta_i}}{RT}\right)$, it follows that

$$-2Ln(\lambda(Y)) = 2RT[Ln(n) - h(\Gamma)]. \tag{21}$$

4 Different symbolizations for different null hypotheses

According to the general framework above stated we focus now on steps (1) and (2) of the general depicted procedure. Given that each hypothesis will require a particular symbolizing map (step 1), in this section we present different symbolization procedures (step 2) to test some interesting nulls.

4.1 Symbolization Map for the null of serial independence

In the case of a time series process, Matilla and Ruiz (2008) used the following symbolizing map to test for serial independence: Let $\{X_t\}_{t\in I}$ be a real-valued time series. For a positive integer

$m \geq 2$ we denote by $\Gamma_1 = S_m$ the symmetric group of order $m!$, that is the group formed by all the permutations of length m. Let $\pi_i = (i_1, i_2, \ldots, i_m) \in S_m$. The positive integer m is usually known as *embedding dimension*.

Now we define an ordinal pattern for a symbol $\pi_i = (i_1, i_2, \ldots, i_m) \in \Gamma_1$ at a given time $t \in I$. To this end we consider that the time series is embedded in an m-dimensional space as follows:

$$X_m(t) = (X_{t+1}, X_{t+2}, \ldots, X_{t+m}) \text{ for } t \in I$$

Then we say that t is of π_i-type if and only if $\pi_i = (i_1, i_2, \ldots, i_m)$ is the unique symbol in the group S_m satisfying the two following conditions:

$$(a) \quad X_{t+i_1} \leq X_{t+i_2} \leq \cdots \leq X_{t+i_m}, \text{ and}$$

$$(b) \quad i_{s-1} < i_s \text{ if } X_{t+i_{s-1}} = X_{t+i_s}$$

Condition (b) guaranties uniqueness of the symbol π_i. This is justified if the values of X_t have a continuous distribution so that equal values are very uncommon, with a theoretical probability of occurrence of 0.

Then we define the symbolization map as $f_1 : \{X_t\}_{t \in I'} \to \Gamma_1$ given by

$$f_1(X_t) = (i_1, i_2, \ldots, i_m) \tag{22}$$

where $(i_1, i_2, \ldots, i_m) \in \Gamma_1$ is such that t is of $(i_1, i_2, \ldots, i_m)$-type.

Moreover, under the null of independence the distribution of the symbols is uniform and therefore the map f_1 is a *standard* symbolization map.

4.2 Symbolization Maps for the null of Spatial independence

In the case of spatial processes, López et al. (2009) give a symbolization procedure to test for spatial independence as follows: Let $\{X_s\}_{s \in S}$ be a real-valued spatial process, where S is a set of coordinates. Given a location s_0, we will denote by (ρ_i^0, θ_i^0) the polar coordinates of location s_i taking as origin s_0.

Let $m \in \mathbb{N}$ with $m \geq 2$. Next, we consider that the spatial process $\{X_s\}_{s \in S}$ is embedded in an m-dimensional space as follows:

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

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

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

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

Notice that conditions (a) and (b) ensure the uniqueness of $X_m(s)$ for all $s \in S$.

The proposed standard symbolization map f is defined as follows: denote by Me the median of the spatial process $\{X_s\}_{s\in S}$ and let

$$\delta_s = \begin{cases} 0 & \text{if } X_s \leq Me \\ 1 & \text{otherwise} \end{cases} \tag{23}$$

Now, define the indicator function

$$\mathcal{I}_{s_1 s_2} = \begin{cases} 0 & \text{if } \delta_{s_1} \neq \delta_{s_2} \\ 1 & \text{otherwise} \end{cases} \tag{24}$$

Then, the standard symbolization map

$$f_2 : \{X_s\}_{s\in S} \to \Gamma_2 \tag{25}$$

is defined as:

$$f_2(X_s) = (\mathcal{I}_{ss_1}, \mathcal{I}_{ss_2}, \ldots, \mathcal{I}_{ss_{m-1}}) \tag{26}$$

Notice that under the null of independence the distribution of the symbols is uniform and therefore the map f_2 is a *standard* symbolization map.

We now consider another possible symbolization procedure for the spatial process $\{X_s\}_{s\in S}$ to test for spatial-independence. Let $\Gamma_3 = \{1, 2, \ldots, k\} \times \{1, 2, \ldots, k\}$. Let N_s be the set of locations s formed by the neighbors of s and let n_s its cardinality. Denote by $X_s^N = \frac{1}{n_s} \sum_{s' \in N_s} X_{s'}$. Denote by q_i and q_i^N the i-th quantile of the variables X and X^N respectively for $i \in \{1, 2, \ldots, k-1\}$. We will denote by $q_0 = \min_{s\in S} X_s$ (resp $q_0^N = \min_{s\in S} X_s^N$) and $q_{k+1} \max_{s\in S} X_s$ (resp. $q_{k+1}^N = \max_{s\in S} X_s^N$). Then we define the symbolization map

$$f_3(X_s) = (i, j) \tag{27}$$

if and only if $X_s \in [q_{i-1}, q_i]$ and $X_s^N \in [q_{j-1}^N, q_j^N]$.

Again, under the null of independence the distribution of the symbols is uniform and therefore the map $f_3 : \{X_s\}_{s\in S'} \to \{1, 2, \dots, k\} \times \{1, 2, \dots, k\}$ is a *standard* symbolization map.

It is also possible to construct a *non-standard* symbolization map to test for independence in the spatial context. The following symbolization is an example that underlines the potential use of non-standard symbolization procedures to model specification analysis.

Consider the set Γ_2 of symbols defined in (26) for a fixed embedding dimension m.

Let us define the following equivalence relation $\sim$:

$$(\mathcal{I}_{ss_1}, \mathcal{I}_{ss_2}, \dots, \mathcal{I}_{ss_{m-1}}) \sim (\mathcal{I}_{s's'_1}, \mathcal{I}_{s's'_2}, \dots, \mathcal{I}_{s's'_{m-1}})$$

if and only if there exists an integer k such that $\mathcal{I}_{s's'_i} = \mathcal{I}_{ss_{\overline{i+k}}}$ for all $i \in \{1, 2, \dots, m-1\}$, where we denote by $\overline{a}$ the remainder of the division of a over $m-1$.

Consider as a set of symbols $\Gamma_4 = \Gamma_2/\sim$, that is, the set of equivalence classes in Γ_2 modulo the equivalence relation $\sim$.

Notice that, in general, in this case not all the symbols in Γ_4 have the same probability to occur and therefore the symbolization map $f_4 : \{X_s\}_{s\in S'} \to \Gamma_4$ is *non-standard*.

4.3 Symbolization Maps for the null of Spatiotemporal Independence

Finally we consider the case of spatiotemporal processes $\{X_{ts}\}_{t\in I', s\in S'}$. In this case as in the previous ones, it is possible to define several standard and non-standard symbolization maps. By simplicity we can adapt the previous symbolizations to the spatiotemporal case as follows:

For a fixed location $s_0 \in S'$ define $\{X_{t(s_0)}\}$ as the time series $\{X_{1s_0}, X_{2s_0}, \dots, X_{p(s_0)}, \dots\}$. Similarly for a fixed period $t_0 \in I'$ we define $\{X_{(t_0)s}\}$ as the spatial process $\{X_{t_0s_1}, X_{t_0s_2}, \dots, X_{t_0s_p}, \dots\}$.

Let $m_1, m_2 \in \mathbb{N}$ with $m_1, m_2 \geq 2$ be the time and space embedding dimensions respectively. Then under this setting we define the following symbolization maps $f_{1i} : \{X_{ts}\}_{t\in I', s\in S'} \to \mathcal{S}_m \times \Gamma_i$ for $i = 2, 3$ and 4 defined by:

$$f_{1i}(X_{ts}) = (f_1(X_{ts}), f_i(X_{ts})) \tag{28}$$

where $f_1 : \{X_{t(s)}\} \to S_m$ and $f_i : \{X_{(t)s}\} \to \Gamma_i$.

Notice that, when testing for spatiotemporal independence, when $i = 2, 3$ the symbolization map f_{1i} is standard while for $i = 4$ is non-standard.

5 Independence tests and a general structural change test

We focus now on steps (3) and (4) of the general procedure presented to test hypothesis with symbolic analysis. Finally, we introduce a test for testing any structural change in a wide sense.

5.1 Testing for Serial independence

Using the symbolization map f_1 and (21) under the null of independence Matilla and Ruiz (2008) constructed the following statistic.

Theorem 5.1 (Matilla and Ruiz (2008))**.** *Let Let $f_1 : \{X_t\}_{t\in I'} \to \Gamma_1$ be the symbolization map defined in (22) with $|I'| = T$. Denote by $h(\Gamma_1)$ the permutation entropy defined in (2). If the time series $\{X_t\}_{t\in I}$ is independent, then the affine transformation of the permutation entropy*

$$G(S_m) = 2T[Ln(m!) - h(S_m)] \tag{29}$$

is asymptotically $\chi^2_{m!-1}$ distributed.

5.2 Spatial independence

Using the symbolization map f_2 and (21) under the null of spatial independence López et al. (2009) prove the following result.

Theorem 5.2 (López et al. (2009))**.** *Let $f_i : \{X_s\}_{s\in S'} \to \Gamma_i$, $i = 2, 3$ be the symbolization maps defined in (25) and (27) with $|S'| = R$. Denote by $h(\Gamma_i)$ the symbolic entropy defined in (2). If the spatial process $\{X_s\}_{s\in S}$ is independent, then the affine transformation of the symbolic entropy*

$$SG(\Gamma_i) = 2R[Ln(n_i) - h(\Gamma_i)] \tag{30}$$

is asymptotically $\chi^2_{d_i}$ distributed where $d_2 = 2^{m-1}$, $d_3 = (k-1)^2 + 2$, $n_2 = 2^{m-1}$ and $n_3 = k^2$.

In the table below we show the size and power simulation results for the $SG(\Gamma_3)$ for $k = 4$. We have considered in this experiment various SAR and SMA spatial processes by simulating over different lattices ($R = 49, 144, 262$ and 400) and three ρs (.1, .5 and .9). Rejection percentajes are reported at 5% level. The Monte Carlo experiment and the results for the symbolization map f_2 are available upon request, they are not reported for sake of space.

The models under study are:

$$\begin{aligned} SAR &: X = (I_n - \rho W)^{-1}\varepsilon, \\ SMA &: X = (I_n + \rho W)\varepsilon \\ \text{being } \varepsilon &\sim iid, N(0,1) \end{aligned}$$

Table 1. Empirical Power and Size for SAR and SMA models for a $SG(\Gamma_3)$

		SIZE		POWER		
		U(0,1)	**N(0,1)**		**Rho**	
	R			0.1	0.5	0.9
	49	0,046	0,043	0,042	0,446	0,988
SAR	144	0,044	0,042	0,068	0,929	1
	262	0,049	0,054	0,095	0,992	1
	400	0,051	0,052	0,136	1	1
	R					
	49	0,046	0,043	0,046	0,386	0,884
SMA	144	0,044	0,042	0,072	0,895	1
	262	0,049	0,054	0,083	0,945	1
	400	0,051	0,052	0,123	1	1

In general, it can be observed that this symbolic based test performs really well in terms of power and size.

5.3 Spatiotemporal Independence

Similarly, one can get a test for independence by using the standard symbolization map f_{1i} for $i = 2, 3$. More concretely we obtain the following result.

Theorem 5.3. *Let $f_{1i} : \{X_{ts}\}_{t\in I', s\in S'} \to S_m \times \Gamma_i$ be the standard symbolization maps defined in (28) with $i = 2, 3$ and 4. Denote by $h(S_m \times \Gamma_i)$ the symbolic entropy defined in (2). If the spatiotemporal process $\{X_{ts}\}_{t\in I, s\in S}$ is independent, then the affine transformation of the symbolic entropy*

$$STG(S_{m_t} \times \Gamma_i) = 2RT[Ln(n_i) - h(S_{m_t} \times \Gamma_i)] \tag{31}$$

is asymptotically $\chi^2_{d_i}$ distributed where $d_2 = (m_t - 1)!2^{m-1}$, $d_3 = (m! - 1)((k-1)^2 + 2)$, $n_2 = m_t!2^{m_s-1}$ and $n_3 = m!k^2$.

In the table below you can fin the size and power of the $STG(S_{m_t}, \Gamma_2)$ (standard) for $m_t = 2, 4$ and $m_s = 4$. Under the alternative, we have considered the following data generating processes

$$y_t = (I - \gamma W)^{-1}(\alpha y_{t-1} + \eta W y_{t-1} + \Delta_t + \varepsilon_t) \tag{32}$$

where $\varepsilon_t \sim iid$, $N(0, 1)$ and the parameters α, γ and η introduce temporal, spatial and spatiotemporal dependence respectively. The parameter Δ_t has been considered constant and equal to 5. The entries marked with an asteristc denotes that simulations are over a non-regular lattice.

Table 4. Empirical Power and Size for SAR and SMA models for $STG(S_{m_t},\Gamma_2)$ Tests

			SIZE				POWER								
		α	0	0.1	0.3	0.5							0.1	0.2	0.3
		γ	0				0.1	0.3	0.5				0.1	0.2	0.3
T/R	m_t	η	0							0.1	0.3	0.5	0.1	0.2	0.3
T=20	2		0.036	0,149	0,121	0,171	0,466	1,000	1,000	0,069	0,131	0,327	0,609	1,000	1,000
R=100	4		0.056	0,185	0,939	1,000	0,206	0,995	1,000	0,072	0,125	0,379	0,428	1,000	1,000
T=30	2		0.034	0,132	0,120	0,174	0,949	1,000	1,000	0,116	0,163	0,697	0,988	1,000	1,000
R=225	4		0.036	0,511	1,000	1,000	0,625	1,000	1,000	0,062	0,249	0,885	0,971	1,000	1,000
T=20	2		0.100	0,103	0,136	0,164	0,609	1,000	1,000	0,107	0,111	0,380	0,474	1,000	1,000
R=259*	4		0.084	0,211	0,969	1,000	0,240	0,999	1,000	0,109	0,146	0,387	0,722	1,000	1,000

* Denotes non-regular lattice. α, γ and η introduce temporal, spatial and spatiotemporal dependence.

Notice that regarding the time independence, the size of the tests is correct with respect to the number of symbols and in regular lattices the size is near the nominal size. Regarding the power, when $m_t = 2$ the results are poor because is not possible to gather any structure with this embedding dimension and the results improve for $m_t = 4$. As regards the spatial and spatiotemporal dependence, the test behave well in terms of power.

5.4 Structural change

In this subsection the term 'structural change' is understood in a broad sense. Particularly, we think of any change that modifies the time and/or the spatial configuration of the data generating process (DGP). Structural changes occur, but we do not know when nor where. For any particular structural change, a symbolization map can be implemented to test for it.

Let $f : \{X_{ts}\}_{s\in S} \to \Gamma$ be a symbolization map. The most general (in terms of structural change) null hypothesis that we want to test is:

H_0: The distribution of $\{X_{(t)s}\}_{s\in S}$ and $\{X_{(t')s}\}_{s\in S}$ are identical for all $t, t' \in I$

Therefore the null can be restated in terms of symbols as follows:

$$H_0 : p_\eta(t) = p_\eta(t') \text{ for all } t, t' \in I \text{ and all } \eta \in \Gamma. \tag{33}$$

Now, under the null, we have that $p_\eta(t) = \frac{1}{T} p_\eta$ for all symbol $\eta \in \Gamma$ and hence we get that

$$\lambda(Y) = (RT)^{RT} \frac{\left(\frac{1}{T}\right)^{\sum_{i=1}^{n}\sum_{t=1}^{T} n_{\eta_i}(t)} \prod_{i=1}^{n} p_{\eta_i}^{\sum_{t=1}^{T} n_{\eta_i}(t)}}{\prod_{i=1}^{n}\prod_{t=1}^{T} n_{\eta_i(t)}^{n_{\eta_i}(t)}} = (RT)^{RT} \frac{\left(\frac{1}{T}\right)^{RT} \prod_{i=1}^{n} p_{\eta_i}^{n_{\eta_i}}}{\prod_{i=1}^{n}\prod_{t=1}^{T} n_{\eta_i(t)}^{n_{\eta_i}(t)}}. \tag{34}$$

On the other hand, $SC(\Gamma) = -2Ln(\lambda(Y))$ asymptotically follows a Chi-squared distribution with $(T-1)n$ degrees of freedom. Hence the estimator $\widehat{SC}(\Gamma)$ of $SC(\Gamma)$ is:

$$\begin{aligned}
\widehat{SC}(\Gamma) &= -2\left[RT\ln(RT) - RT\ln(T) + \sum_{i=1}^{n} n_{\eta_i}\ln(\frac{n_{\eta_i}}{RT}) - \sum_{t=1}^{T}\sum_{i=1}^{n} n_{\eta_i}(t)\ln(n_{\eta_i}(t))\right] \\
&= -2RT\left[\ln(RT) - \ln(T) + \sum_{i=1}^{n} \frac{n_{\eta_i}}{RT}\ln\left(\frac{n_{\eta_i}}{RT}\right) - \sum_{t=1}^{T}\sum_{i=1}^{n} \frac{n_{\eta_i(t)}}{RT}\ln(n_{\eta_i}(t))\right] \\
&= 2RT\left[\ln(T) - \sum_{i=1}^{n} \frac{n_{\eta_i}}{RT}\ln\left(\frac{n_{\eta_i}}{RT}\right) + \sum_{t=1}^{T}\sum_{i=1}^{n} \frac{n_{\eta_i(t)}}{RT}\ln\left(\frac{n_{\eta_i}(t)}{RT}\right)\right] \\
&= 2RT\left[\ln(T) + \sum_{t=1}^{T} \widehat{h}_t(\Gamma) - \widehat{h}(\Gamma)\right]
\end{aligned} \tag{35}$$

And hence we have proved the following result.

Theorem 5.4. *Let $\{X_{ts}\}_{t\in I, s\in S}$ be a spatiotemporal process. Let $\Gamma = \{\eta_1, \eta_2, \ldots, \eta_n\}$ be a set of n symbols used for the symbolization of the series $\{X_{(t)s}\}_{s\in S'}$ and $f : \{X_{ts}\}_{s\in S} \to \Gamma$ be a symbolization map. If $\{X_{(t)s}\}_{s\in S'}$ and $\{X_{(t')s}\}_{s\in S'}$ are identical for all $t, t' \in I$ then*

$$SC(\Gamma) = 2RT[\ln(T) + \sum_{t=1}^{T} h_t(\Gamma) - h(\Gamma)]$$

is asymptotically $\chi^2_{n(T-1)}$.

This theorem holds for any particular symbolization map. In the interesting case of temporal break in a spatial process, the general test can be implemented to test the null of no parameter

break. To show the convenience of this test we simulate below two processes for $T = 2$ and $T = 4$. Also the spatial structure is for $m_s = 4$ (each location is related with three neighbors). Sample sizes are $R = 49, 100, 400$ and 900. We also consider different sets of breaks by varying ρ_t.

We have utilized the map $f_{1i} : \{X_{ts}\}_{t\in I', s\in S'} \to S_m \times \Gamma_2$ as the standard symbolization map defined in terms of (28).

The DGP's used in the simulations are:

$$DGP1 : X_t = (I - \rho_t W)^{-1}\varepsilon_t$$
$$DGP2 : X_t = (I + \rho_t W)^{-1}\varepsilon_t$$

with $\varepsilon_t = N(0, 1)$.

Table 5. *Size* and Power for the Structural Change Test for SAR

ρ_1	ρ_2	ρ_3	ρ_4	49	100	900
0.0	*0.0*	-	-	*0.048*	*0.040*	*0.040*
0.5	*0.5*	-	-	*0.057*	*0.054*	*0.061*
0.8	*0.8*	-	-	*0.085*	*0.092*	*0.085*
0	0.3	-	-	0.087	0.098	0.754
0	0.5	-	-	0.178	0.331	0.999
0	0.8	-	-	0.480	0.819	1.000
0.2	-0.3	-	-	0.161	0.297	0.996
0.0	*0.0*	*0.0*	*0.0*		*0.043*	*0.028*
0.5	*0.5*	*0.5*	*0.5*		*0.058*	*0.045*
0.8	*0.8*	*0.8*	*0.8*		*0.157*	*0.137*
0.0	0.1	0.2	0.3		0.075	0.592
0.0	0.3	0.6	0.9		0.951	1.000
0.1	0.5	0.1	0.5		0.261	0.998
-0.3	0.1	0.4	0.0		0.299	1.000

Table 6. *Size* and Power of the Structural Change for SMA process

ρ_1	ρ_2	ρ_3	ρ_4	49	100	900
0.0	*0.0*	-	-	*0.058*	*0.040*	*0.041*
0.5	*0.5*	-	-	*0.051*	*0.042*	*0.041*
0.8	*0.8*	-	-	*0.051*	*0.049*	*0.060*
0	0.3	-	-	0.084	0.104	0.718
0	0.5	-	-	0.129	0.213	0.991
0	0.8	-	-	0.280	0.500	1.000
0.2	-0.3	-	-	0.156	0.276	0.993
0.0	*0.0*	*0.0*	*0.0*		*0.033*	*0.030*
0.5	*0.5*	*0.5*	*0.5*		*0.045*	*0.029*
0.8	*0.8*	*0.8*	*0.8*		*0.054*	*0.048*
0.0	0.1	0.2	0.3		0.088	0.547
0.0	0.3	0.6	0.9		0.405	1.000
0.1	0.5	0.1	0.5		0.179	0.978
-0.3	0.1	0.4	0.0		0.289	0.999

From both tables it can be observed that the power of the test increases as more spatial and/or time data are available, while the size of the test remains controled.

6 Conclusions

In this paper we have shown the flexibility and the power of the symbolic permutation entropy approach in order to deal with several types of variables: time series, spatial series or panel data. We offered a unified non-parametric framework in which it is possible to solve different inference problems related to the stochastic structure of the data. The method is, statistically and computationally, simple but powerful and very competitive against other well-established procedures proposed in the literature.

The key aspect of our procedure is to define correctly the set of symbols more appropriate to tackle with the null hypothesis under analysis. In each case, there will exist various alternatives, standard or non-standard, and the problem is to make a good choice among them.

7 References

Arellano, M (2003): Panel Data Econometrics. Oxford: Oxford University Press.

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

Hsiao, Ch. (2003): Analysis of Panel Data (2nd edition). Cambridge: Cambridge University Press.

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.

López, F., M. Matilla, J. Mur and M. Ruiz (2009): A Non-Parametric Spatial Independence Test Using Symbolic Entropy. Working Paper, Politechical University of Cartagena.

Maasoumi, E. (1993): A Compendium to Information Theory in Economics and Econometrics. Econometric Reviews, 12, 137-181.

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

Nerlove, M. (2002): Essays in Panel Data Econometrics. Cambridge: Cambridge University Press.

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

Ullah, A (1993): Entropy, Divergence and Distance Measures with Econometric Applications. Journal of Statistical Planning and Inference, 49, 137-162.