

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Symbolic Time Series Analysis and Its Application in Social Sciences

Wiston Adrián Risso

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.70826>

Abstract

The present chapter intends to present the symbolic time series analysis (STSA) reviewing the recent developments in sciences. Even if there are very few works applied to social sciences, STSA has a potential to be developed. In particular, due to the limitations about historical data, fields such as Economics and Finance need to develop statistical tests to prove their hypotheses. An independence test and a causality test based on STSA are reviewed. They seem to be more powerful, detecting different kinds of nonlinearities compared with the classical tests, usually applied in social sciences. However, there is much work to do with STSA, and social sciences are a fertile field for the development of new powerful tools.

Keywords: STSA, information theory, entropy, independence, causality

1. Introduction

Time series has a long history in social sciences, especially in economics and finance. As it is well known, much of economics and finances are concerned with modeling dynamics, and systematization of data over time was a subject that appeared early. In particular, two empirical topics become important when working with time series in social sciences: inferences and forecasting. The cumulated historical data permitted to applied statistical methods in order to find evidence of causation between social variables, finding some support to social theories. Considering the nonexperimental nature of the social sciences, this also encourages the development of statistical techniques. In fact, while in physics, it is relatively easy to get hundreds of thousands of data for a given time series, in economics there are often only 50 or 100 data for a time series, and maybe we can obtain thousands of data in financial series. For this reason, much of the statistical effort, in particular econometric effort was focused on developing powerful statistical tests, considering the availability of small samples. This is an important different approach between econometrics and for example, statistical mechanics in theoretical physics.

We can identify two main groups in time series econometrics: univariate time series analysis concerning with techniques for the analysis of dependence in adjacent observations. It has increased importance since 1970 based on the main ideas underlying in [1]; multivariate time series analysis based on the vector autoregressive (VAR) models, made popular by [2]. In the first group, we find all the autoregressive integrated moving average (ARIMA) models and the related generalized autoregressive conditional heteroscedasticity (GARCH) models developed by [3]. The second group is a generalization of the AR models and we can find two important developments based on this: cointegration proposed by [4] focusing on finding a statistical relationship between variables; and noncausality test developed by [5], which takes the concept of predetermination try to test if a variable causes another. Much of the development in time series econometrics is found in books such as [6–18].

In summary, dependence and causation are two important topics in time series econometrics and time series analysis. These topics are related with the importance of inference and forecasting in social sciences. Econometrics has been focused in developing powerful test considering the available small samples. Most of these developments are based on linear models even if there are some developments considering nonlinearities; see for instance [19, 20].

Time series analysis in econometrics is mostly based on observations belonging to the set of the real numbers. Some variables can be categorical such as dummy variables. However, in this chapter, we will talk about a different approach that is known as symbolic time series analysis (STSA). It has been originally applied to physics and engineering as a statistical methodology to detect the very dynamic of highly noise time series. The application to social sciences such as economics or finance is very recent and there are some novel developments.

As mentioned before, the application of STSA in social sciences requires a different approach due to data limitation. In this sense, the design of powerful test considering the availability of data is crucial. As abovementioned, dependence and causation are two important topics. In this sense, we review an independence test and a first approach on testing noncausality, both based on STSA. The information theory was adopted as an approach to analyze the symbolic time series and the approximation of Shannon Entropy as an important measure, applied to test design.

The chapter is organized as follows. Section 2 presents the symbolic time series approach and its relation with the symbolic dynamics. In Section 3, we review some of the literature of STSA applied to the sciences. In Section 4, the information theory approach and Shannon Entropy measure is explained. Section 5 presents a review of the independence symbolic test. Section 6 focuses on causality test based on STSA. Section 7 discusses the difference between the proposed symbolic noncausality test and the traditional and well-known Granger noncausality test. Finally, in Section 8, we draw some conclusions and present some future lines of research.

2. Symbolic time series analysis

The concept of symbolization has its roots in dynamical systems theory, particularly in the study of nonlinear systems, which can exhibit bifurcation and chaos. In [21], it is asserted that

symbolic dynamics is a method for studying nonlinear discrete-time systems by taking a previously codified trajectory using strings of symbols from a finite set, also called an alphabet. According to [22], symbolic dynamics and symbolic analysis are connected but are different concepts. In fact, the former is the practice of modeling a dynamical system by a discrete space. However, the latter is an empirical approach to characterize highly noisy data by considering a partition, discretizing the data, and obtaining a string representing the very dynamic of the process.

As asserted by [23], symbolization involves transformation of raw time series measurements into a series of discretized symbols that are processed to extract information about the generating process. In this way, we can search for nonrandom patterns and dependence by transforming a given time series $\{x_1, x_2, \dots, x_T\}$ into a symbolic string $\{s_1, s_2, \dots, s_T\}$.

The STSA approach is easy to apply but the definition of the right partition is the most difficult thing to do. Generally, it applied an equiprobable partition implying to take the empirical distribution of a given time series $\{x_1, x_2, \dots, x_T\}$ and establishing two or more equally probable regions. For instance, for a Gaussian time series, we can define two equally probable regions considering as partition the mean equal to zero. After that, we can assign the symbol $s_i = 0$ for negative values and $s_i = 1$ for positive ones. In this way, we transform a continuously random series into a discrete string similar to the outcomes from flipping a coin.

3. STSA in applied sciences

In [23], the applications of STSA techniques to the different fields of science are reviewed. According to the authors, the different applications suggest that symbolization can increase the efficiency of finding and quantifying information from the systems. Mechanical systems were one of the first applications where symbolic analysis was successfully used to characterize complex dynamics. In [24–26], symbolic methods to the analysis of experimental combustion data from internal combustion engines are applied. Their objective was to study the onset of combustion instabilities as the fueling mixture was leaned. STSA has also been applied in Astrophysics and Geophysics. For instance, [27] analyzes weak-reflected radar signals from the planet Venus to measure the rotational period. In [28], a binary symbolization to analyze solar flare events is utilized. Biology and Medicine is another field where STSA has been applied. There have been many recent applications of symbolic analysis for biological systems, most notably for laboratory measurements of neural systems and clinical diagnosis of neural pathologies. STSA has been applied in neurosciences. In [29, 30], symbolization data is applied to equal-sized interval to partition EEG signals to identify seizure precursors in electroencephalograms. [31] proposed a new damage localization method based on STSA to detect and localize a gradually evolving deterioration in the system. They assert that this method could be demanded for implementation in real-time observation application such as structural health monitoring. In [32], the STSA is used in human gait dynamics. The results of this study can have implication modeling physiological control mechanism and for quantifying human gait dynamics in physiological and stressed conditions. In [33], the heart-rate dynamics is studied by using partitions aligned on the data mean and ± 1 and ± 2 sample standard deviations, for a

symbol-set size of 6. In [34], the prevalence of irreversibility in human heartbeat is analyzed applying STSA.

Application of symbolization to fluid flow measurements has spanned a wide range of data types from global measurements of flow and pressure drop, to formation and coalescence of bubbles and drops, to spatiotemporal measurements of turbulence. In [35], an approach for transforming images of complex flow fields (as well as other textured fields) into a symbolic representation is developed. In [36], STSA is applied to the networks of genes, which is important underlying the normal development and function of organisms. Information about the structure of the genome of humans and other organisms is increasing exponentially. In [37], equiprobable symbols are used for analyzing measurements from free liquid jets in order to readily discriminate between random and nonrandom behavior. In [38], STSA is applied to the detection of incipient fault in commercial aircraft gas turbine engines. In [39], combustion instability in a swirl-stabilized combustor is investigated using STSA. Chemistry-related applications of symbolic techniques have been developed for chemical systems involving spontaneous oscillations or propagating reaction fronts. In [40], a type of symbolization for improving the performance of Fourier-transform ion-cyclotron mass spectrometry is applied. Artificial Intelligence, Control, and Communication are fields where symbolization has been incorporated. In [41], a phase-space partitioning to model communication is used. An example application of symbolization to communication is found in [42], utilizing small perturbations to encode messages in oscillations of the Belousov-Zhabotinsky (BZ) reaction. In robotics, a symbolic time series-based statistical learning method to construct the generative models of the gaits (i.e., the modes of walking) for a robot, see [43], has been developed. Efficacy of the proposed algorithm is demonstrated by laboratory experimentation to model and then infer the hidden dynamics of different gaits for the T-hex walking robot. In [44], an algorithm to intuitively cluster groups of agent trails from networks based on STSA is proposed. The authors assert that temporal trails generated by agents traveling to various locations at different time epochs are becoming more prevalent in large social networks. The algorithm was applied to real world network trails obtained from merchant marine ships GPS locations. It is able to intuitively detect and extract the underlying patterns in the trails and form clusters of similar trails.

The methods of data symbolization have also been applied for data mining, classification, and rule discovery. In [45], rule discovery techniques to real-valued time series via a process of symbolization are applied. Finally, we find some applications of STSA in Social Science. In [46–48], STSA and minimal spanning tree (MST) are applied to construct cluster of financial asset with application to portfolio theory. Utilizing a similar methodology, in [49], the dynamics of exchange market is studied, and in [50], the international hotel industry in Spain is analyzed. In [51, 52], STSA and entropy are applied to measure informational efficiency in financial markets.

4. Information theory and Shannon entropy

The term entropy was first used by Rudolf Clausius in [53] related to the second law of thermodynamics. Subsequently, the communication theory [54] used the Shannon entropy as

a measure of uncertainty where the maximum entropy corresponds to the maximum degree of uncertainty. In this sense, a random process will take the maximum entropy value. In fact, English language is not a random process; some patterns such as "THE" are more probable than sequences such as "DXC". Note, that in a random process, the two sequences should have the same probability. This principle is very relevant because if a symbolic string is random, the entropy should be the maximum.

The entropy measure (H) must meet the following conditions:

1. $H(P)$ should be a function of the probability distribution of the n events expressed as the vector $P = (p_1, p_2, \dots, p_n)$.
2. (Continuity), $H(P)$ should be a continuous function of vector P .
3. (Symmetry), the measure should be unchanged if the outcomes p_i are re-ordered.
4. (Expansible), Event of probability zero should not contribute to the entropy, $H(p_1, p_2, \dots, p_n, 0) = H(p_1, p_2, \dots, p_n)$.
5. (Minimum), the measure should take value 0 when there is not uncertainty.
6. (Maximum), the measure should be maximal if all the outcomes are equally likely. It means $p_1 = p_2 = \dots = p_n = 1/n$.
7. For equiprobable events, the entropy increases with the number of outcomes. $H(p_1 = 1/(n + 1), \dots, p_{n+1} = 1/(n + 1)) > H(p_1 = 1/n, \dots, p_n = 1/n)$.

In [54], the Shannon entropy function is proposed:

$$H_n(P) = - \sum p_i \log_2(p_i) \tag{1}$$

The entropy is frequently measured in bits by using log base 2 satisfying all the properties already mentioned. Note that the maximum property is confirmed solving the following Lagrangian expression (2).

$$- \sum p_i \log_2(p_i) - \lambda \left(\sum p_i = 1 \right) \tag{2}$$

The Shannon entropy is concaved with a global maximum when all the probabilities are equal. In addition, when $p_i = 0$, the convention that $0 \cdot \log 0 = 0$ is used. Thus, adding zero, probability terms do not change the entropy value.

In order to clarify the concept of Shannon, consider two possible events and their respective probabilities p and $q = 1 - p$. The Shannon entropy will be defined by Eq. (3).

$$H = -(p \cdot \log(p) + q \cdot \log(q)) \tag{3}$$

Figure 1 shows graphically the function shape, note that the maximum is obtained when the probability is 0.5 for each event. This case corresponds to a random event; on the other hand, note that a certain event (when probability of one event is 1) will produce entropy equal to 0.

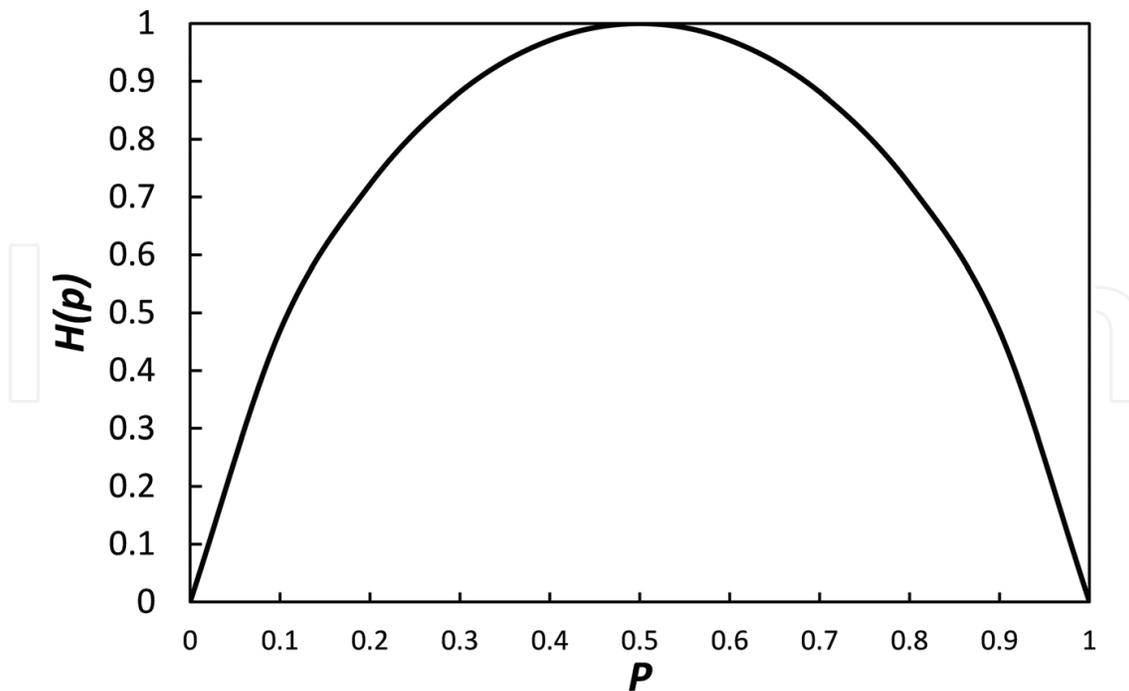


Figure 1. Shape of the Shannon entropy function. Note that maximum happens when the process is random ($p = 0.5$).

In general, [55] showed that any measure satisfying all the properties must take the following form:

$$-c \sum p_i \log_2(p_i) \quad (4)$$

In order to normalize the Shannon entropy, c usually takes the value $1/\log_2(n)$ allowing to compare events of different sizes.

5. Symbolic independence test

STSA seems to present a good performance when detecting independence in time series. A variety of dynamical processes are present in economics. Linearity, nonlinearity, deterministic chaos, and stochastic models have been applied when modeling a complex reality. In [56], a runs test is designed, asserting that the problem of testing randomness arises frequently in quality control of manufactured products. It is remarked that detecting dependence in time series is an essential task for econometricians and applied economist. In [57], the well-known BDS test is introduced, considered as a powerful test to detect nonlinearity. In [58], a simple and powerful test based on STSA is proposed and the results are compared with the BDS and runs test. On one hand, it is found that BDS is not able to detect processes such as the chaotic Anosov and the stochastic processes nonlinear sign model (NLSIGN), nonlinear autoregressive model (NLAR), and nonlinear moving average model (NLMA). On the other hand, runs test cannot detect the chaotic Anosov, the logistic process, the bilinear, the NLAR, and the NLMA stochastic processes. The experiments show that the test based on STSA has no problem

detecting all these dynamics. It is concluded that proposed test is simple, easy to compute, and is powerful with respect to the other two tests. In particular, for small samples, it is the only one able to detect models such as chaotic Anosov and nonlinear moving average (NLMA). Besides, the test is applied to financial time series to detect nonlinearity on the residuals after applying a GARCH model. In this case, the BDS rejected the independence few times whereas the SRS test still detects nonlinearity in the residuals. It seems that BDS considers that the GARCH(1,1) model is a good model most of the time. However, the symbolic test suggests that GARCH(1,1) would not be a good model considering all the nonlinear components.

Here, we review briefly the test and repeat some experiments comparing the results with the well-known BDS and runs tests. At first, let us consider a finite time series generated by an independent or random process-sized T^* $\{x_t\}_{t=1,2,\dots,T^*}$. Define a partition in the series in "a" equiprobable regions obtaining the symbolized time series $\{s_t\}_{t=1,2,\dots,T^*}$, where each symbol s_t takes a symbolic value from the alphabet $A = \{A_1, A_2, \dots, A_a\}$. Since, we want to derive a general statistic for different alphabet sizes a and different subsequences lengths w , we have to make two considerations: (1) from now, we will call n to the quantity of possible events. That is $n = a^w$, where for the simplest case ($w = 1$) implies $n = a$, then the quantity of events is equal to the symbol-set size; (2) in practice, we have a finite sample size T^* , there is no problem for $w = 1$, but when we compute subsequences or time-windows w of consecutive symbols we loss observations. For example, when we compute the frequency for two consecutive symbols, we have a total sample size $T^* - 1$. In general, we can define the sample size $T = T^* + w - 1$, again for the trivial case $w = 1$, $T^* = T$

Note that defining S_i for $i = 1, 2, \dots, n$ as the sum of the total i events in the time series, we can derive the multidimensional variable $S = \{S_i/T\}$ being distributed as a multinomial with $E(S_i/T) = (1/n)$, $Var(S_i/T) = (1/n)(n-1)/nT$ and $Cov(S_i/T, S_j/T) = -(1/n)(1/nT) \forall i \neq j$. As we will see, frequencies of the events should be important in the statistic and the vector of the n frequencies S_i/T could be approximated by a multivariate normal distribution $N(1/n, \sigma^2 \Sigma)$ where σ^2 is $(1/nT)$ and Σ is a idempotent matrix as in (5)

$$\Sigma_{n \times n} \equiv \begin{bmatrix} (n-1)/n & -1/n & \dots & -1/n \\ -1/n & (n-1)/n & \dots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \dots & (n-1)/n \end{bmatrix} \quad (5)$$

For convenience, we can define the normalized vector variable $\{\varepsilon_i\} = \{(S_i/T) - (1/n)\}_{i=1,2,\dots,n}$ having a multivariate normal distribution $N(\emptyset, \sigma^2 \Sigma)$, being \emptyset , the null vector. Then, the statistic can be defined as a quadratic form in random normal variables (6).

$$\left\{ \frac{\sum_{i=1}^{i=n} \varepsilon_i^2}{\sigma^2} \right\} \quad (6)$$

In [58] is applied the distribution of quadratic forms in normal variables presented in [59]. $X = (\varepsilon_1/\sigma, \varepsilon_2/\sigma, \dots, \varepsilon_n/\sigma)$ is distributed multivariate normal $N(\emptyset, \Sigma)$. The theorem indicates that

$tr(A\Sigma) = n-1$, and thus $X'AX$ distributes Chi-square with $(n-1)$ degrees of freedom. In this case, A is the identity matrix I , and Σ is symmetric, singular, and idempotent. Remembering that $\sigma^2 = (1/nT)$, then we obtain that the distribution of the symbolic randomness statistic (SRS) as in (7).

$$SRS \equiv Tn \left\{ \sum_{i=1}^{i=n} \left(\frac{S_i}{T} - \frac{1}{n} \right)^2 \right\} \text{ asymptotically distributes } \chi_{n-1}^2 \quad (7)$$

Note that in practice computing the statistic is very simple. We just have to consider the symbols (a) and subsequences or length (w) and compute the frequencies for each event ($n = a^w$) in the time series.

The algorithm to compute the test is as follows:

Step 1: Considering time series $\{x_t\}_{t=1,2,\dots,T^*}$, compute the empirical distribution, and define equiprobable regions according to the quantity of symbols or the alphabet size.

Step 2: According to the partition, translate $\{x_t\}_{t=1,2,\dots,T^*}$ into $\{s_t\}_{t=1,2,\dots,T^*}$, the symbolic time series when $w = 1$.

Step 3: Compute different symbolic time series for different lengths w , remember that the obtained series in step 2 corresponds to $w = 1$.

Step 4: For each w , compute the frequency of the n different events S_i/T for $i = 1, 2, \dots, n$.

Step 5: For each w , compute the $SRS(a,w) = Tn\{\Sigma(S_i/T - 1/n)^2\}$ as shown in Eq. (7).

Step 6: Compare the $SRS(a,w)$ with the Chi-2 with $n-1$ degree of freedom at 0.05 of significance, under the independence null hypothesis. When $SRS(a,w)$ is larger than the critical value we reject the null hypothesis.

In [58], it is found that the statistic introduced in (7) is related to the Shannon entropy (H). We can derive the approximation expressed in Eq. (8).

$$SRS \approx (1 - H) \cdot T \cdot \ln(n) \quad (8)$$

Note the generalization implied in STSA permits to study different dynamical process. For instance, consider a string of the first 3000 letters from the book "A Christmas Carol", $s_1 = \{\text{marleywasdeadtobeginwith...scroogecar}\}$ and a random string of 3000 letters from an alphabet of 26, $s_2 = \{\text{iskynbmlhjp...vbbihjfk}\}$. Imagine testing this kind of process with BDS or runs test. However, note that would be easy to test this dynamics with the symbolic test. In this case, we can define an alphabet of 26 letters and the string. On the one hand, applying the $SRS(26,1)$ and $SRS(26,2)$ for the s_1 , we obtain the following values 2102.40 and 12331.26, respectively. On the other hand, $SRS(26,1)$ and $SRS(26,2)$ for the string s_2 are 25.79 and 690.26, respectively. Considering that a Chi-2 with 25 degree of freedom at 95% is 37.65 and a Chi-2 with 675 degree of freedom (26^2-1) at 95% is 736.55. Since, the statistics for s_1 are large than the critical value, we can conclude that the process is not random. However, since the statistics for s_2 are less than critical values, we cannot reject the hypothesis of independence.

In [58] is shown that the test is conservative, rejecting the null hypothesis less time than expected. However, it is powerful in detecting nonrandom and nonlinear processes. Considering the four sample sizes, selecting two symbols and length 4 presents decent results in most of the cases. Selecting three symbols seems to be a relative good option for size of 200 or larger and three symbols for a sample size of 500 or larger. The best result is given for a sample of 2000 applying three symbols and length 4. **Table 1** presents the experiments using 1000 Monte Carlo simulations on Normal, Logistic, NLMA, Anosov, and NLSIGN processes reproducing the experiments in [58].

Note that the symbolic test is more conservative than BDS and Runs test when rejecting independence in a normal random process. However, the symbolic test is powerful in detecting nonlinearities in the studied processes. For a sample of 50, Logistic model is detected 100% by the symbolic test, but BDS detects 68%, and Runs test rejects independence 23.90% of the time. Logistic model is still hard to be detected by the run test when sample increases to 2000. Note that NLMA model is detected by the symbolic test when sample is 500 or larger, but it is not detected by BDS and Runs test. It is interesting to note that the chaotic process of Anosov is detected by the symbolic test for a sample larger than 500 but both BDS and Runs tests reject independence less than 6% of the cases. NLSIGN is hard to be detected, for a sample of 2000 the symbolic test detects more than 90% of the cases and Runs test detects 84% of the cases. However, BDS cannot detect the NLSIGN process. In [58] similar results are obtained, the proposed SRS is the only one that is able to detect chaotic Anosov and nonlinear process NLMA when $T = 2000$.

Sample size	Test	Normal (%)	Logistic (%)	NLMA (%)	Anosov (%)	NLSIGN (%)
T = 50	SRS(2,3)	1.20	41.00	1.30	2.90	0.20
	SRS(3,2)	0.70	100.00	0.40	0.80	0.50
	BDS	9.70	68.10	7.60	18.10	12.00
	RUN test	2.90	23.90	1.30	3.90	2.20
T = 500	SRS(2,3)	2.10	19.30	13.50	2.30	9.90
	SRS(3,2)	0.40	100.00	1.70	0.80	1.10
	SRS(4,3)	3.20	100.00	97.50	93.80	17.40
	BDS	3.60	66.40	7.20	5.30	4.40
	RUN test	3.80	14.70	16.30	5.50	24.10
T = 2000	SRS(2,3)	1.90	14.40	62.90	2.50	64.80
	SRS(3,2)	0.30	100.00	25.30	1.00	3.80
	SRS(4,3)	2.50	100.00	100.00	100.00	92.30
	SRS(5,3)	1.70	100.00	100.00	100.00	92.20
	BDS	2.80	80.40	14.60	3.60	4.00
	RUN test	4.40	12.20	50.00	5.70	84.30

Table 1. Simulated size of the SRS, Runs and BDS statistics.

6. Symbolic noncausality test

The present section reviews the symbolic noncausality test (SNC) and discusses the differences with the classical Granger noncausality test. As in the case of independence test, the main idea here is to derive the asymptotic distribution for the statistic when there is no causality between the series. A full explanation of the test is shown in [60].

Let us consider that X and Y are two independent random time series sized $T + 1$ and the symbolized time series can be expressed as $Sx = \{sx_1, sx_2, \dots, sx_{T+1}\}$ and $Sy = \{sy_1, sy_2, \dots, sy_{T+1}\}$. To test causality, we have to define two new series, grouping Sx and Sy in the following way:

$$(1) Sxy = \{(sx_1, sy_2), (sx_2, sy_3), \dots, (sx_{t-1}, sy_t), \dots, (sx_T, sy_{T+1})\}$$

$$(2) Syx = \{(sx_1, sy_2), (sx_2, sy_3), \dots, (sx_{t-1}, sy_t), \dots, (sx_T, sy_{T+1})\}$$

If the alphabet is composed by three symbols, the combination (sx_{t-1}, sy_t) takes a value from the set of nine possible events $\{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$. Note that each event should be independent with probability $1/9$ (Sx and Sy are random). Only if at least one event were deviated from $1/9$, would there be evidence of noncausality.

An alphabet of $a = 3$ symbols determines $n = 3^2 = 9$ possible events in the set of pairs $\{(x_{t-1}, y_t)\}$ or $\{(y_{t-1}, x_t)\}$. Considering “ a ” symbols and the events $n = a^2$, the vector of the n frequencies Exy_i/T and Eyx_i/T could be approximated by a multivariate normal distribution $N(1/n, \sigma^2 \Omega)$ where σ^2 is $(1/nT)$ and Ω is a idempotent matrix as in (9).

$$\Omega_{n \times n} \equiv \begin{bmatrix} (n-1)/n & -1/n & \dots & -1/n \\ -1/n & (n-1)/n & \dots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \dots & (n-1)/n \end{bmatrix} \tag{9}$$

Following a similar approach as in Section 5, the statistics for the both hypothesis can be defined as in (10) and (11).

$$\left\{ \frac{\sum_{i=1}^{i=n} \epsilon x y_i^2}{\sigma^2} \right\} \tag{10}$$

$$\left\{ \frac{\sum_{i=1}^{i=n} \epsilon y x_i^2}{\sigma^2} \right\} \tag{11}$$

The term in brackets in (10), (11) are quadratic forms in random normal variables. Applying the theorem presented in [59], in the present case where vector $X = (\epsilon_1/\sigma, \epsilon_2/\sigma, \dots, \epsilon_n/\sigma)$ is distributed multivariate normal $N(\emptyset, \Omega)$. As mentioned in Section 5, $tr(A\Omega) = n-1$, thus $X'AX$ distributes Chi-square with $(n-1)$ degrees of freedom. In this case, A is the identity matrix I and Ω is symmetric, singular, and idempotent.

Note that we derive the test assuming that X and Y are random processes. However, we can apply the test for stationary time series and optionally apply an autoregressive process if we want to remove linear dependence and testing the noncausality between the residuals of the two series.

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + ux_t \quad (12)$$

$$y_t = \beta_0 + \beta_1 y_{t-1} + uy_t \quad (13)$$

Finally, the statistics of noncausality $SNC(X \rightarrow Y)$ and $SNC(Y \rightarrow X)$ are defined as in (14) and (15).

$$SNC(X \rightarrow Y) \equiv nT \left\{ \sum_{i=1}^{i=n} \left(\frac{Exy_i}{T} - \frac{1}{n} \right)^2 \right\} \text{ asymptotically distributes } \chi_{n-1}^2 \quad (14)$$

$$SNC(Y \rightarrow X) \equiv nT \left\{ \sum_{i=1}^{i=n} \left(\frac{Eyx_i}{T} - \frac{1}{n} \right)^2 \right\} \text{ asymptotically distributes } \chi_{n-1}^2 \quad (15)$$

Note that in practice, computing the statistic is very simple. In summary, the test works as follows:

Step 1: Consider time series $\{x_t\}_{t=1,2,\dots,T+2}$ and $\{y_t\}_{t=1,2,\dots,T+2}$ we can optionally apply an AR (1) to both series as in (12) and (13) in order to eliminate autocorrelation and define the new residuals time series $\{ux_t\}_{t=1,2,\dots,T+1}$ and $\{uy_t\}_{t=1,2,\dots,T+1}$. Note that 1 observation is lost after applying AR(1).

Step 2: In $\{ux_t\}_{t=1,2,\dots,T+1}$ and $\{uy_t\}_{t=1,2,\dots,T+1}$ apply a partition in “ a ” equiprobable regions and translate the series into $\{sx_t\}_{t=1,2,\dots,T+1}$ and $\{sy_t\}_{t=1,2,\dots,T+1}$.

Step 3: According to the two hypothesis, $X \rightarrow Y$ and $Y \rightarrow X$ define the two sets $Sxy = \{(sx_1, sy_2), (sx_2, sy_3), \dots, (sx_{t-1}, sy_t), \dots, (sx_T, sy_{T+1})\}$ and $Syx = \{(sx_1, sy_2), (sx_2, sy_3), \dots, (sx_{t-1}, sy_t), \dots, (sx_T, sy_{T+1})\}$.

Step 4: For Sxy and Syx , compute the frequency of the $n = a^2$ different events Exy_i/T and Eyx_i/T considering $i = 1, 2, \dots, a^2$.

Step 5: Taking into account Eqs. (14) and (15) compute the $SNC(X \rightarrow Y) = nT \{ \Sigma [(Exy_i/T) - (1/n)]^2 \}$ and $SNC(Y \rightarrow X) = nT \{ \Sigma [(Eyx_i/T) - (1/n)]^2 \}$.

Step 6: Finally, two null hypotheses must be contrasted: X does not cause Y , and Y does not cause X . In the first case $SNC(X \rightarrow Y)$ should be compared with a Chi-2 with $n-1$ degree of freedom at 0.05 of significance, if $SNC(X \rightarrow Y)$ is larger than the critical value the null hypothesis is rejected. The same should be done with $SNC(Y \rightarrow X)$.

7. Symbolic noncausality and Granger noncausality

The concept of causality into the experimental practice is due to Clive Granger. The classical approach of Granger causality is based on temporal properties. Although the principle was

formulated for wide classes of systems, the autoregressive modeling framework proposed by Granger was basically a linear model, and as mentioned in [61] the choice was made due to practical reasons. Granger noncausality test is among the most applied tool testing causality. Three limitations should be noted: (1) the classical test has a good performance when the process is linear. This is because it is based on the vector autoregressive model (VAR); (2) there are extension of the classical test to consider nonlinear causality but they are related with a particular nonlinear model; (3) some authors assert that empirical time series are generally contaminated with noise producing what is known as spurious causality or not allowing to detect the causality.

SCN test presented in [60] is a nonparametric noncausality test based on the symbolic time series analysis. The idea is to develop a complementary test to the Granger noncausality, showing strengths in the points where the Granger test is weak. In this sense, the proposed SNC test performs well detecting nonlinear processes, in particular the chaotic processes. In addition, the mentioned problem related with spurious causality should be alleviated. In fact, according to some experiments nonlinear models such as NLAR model, Lorenz map, and models with exponential terms are not detected by Granger test but the SNC identifies these processes. The test is based on information theory considering an approximation of the entropy as the measure of uncertainty of a random variable. Information theory is considered to be a subset of communication theory. However, in [62] is consider that it is much more. It has fundamental contributions to make in statistical physics, computer science, and statistical inference, and in probability and statistics. It is important to highlight and is an important idea relating symbolic analysis, information theory, and the concept of noise. Information theory considers that communication between A and B is a physical process in an imperfect ambient contaminated by noise. Another important concept is the discrete channel, defined as a system consisting of an input alphabet X and output alphabet Y and a probability transition matrix $p(y|x)$ that expresses the probability of observing the output symbol y given that we send the symbol x .

To compare the performance between the classical Granger noncausality and the proposed SNC test, the following stochastic and deterministic models were simulated:

1. AR(1). We consider two independent series generated by autoregressive (AR) processes: $X_t = 0.2 + 0.45X_{t-1} + \varepsilon_{1t}$ and $Y_t = 0.8 + 0.5Y_{t-1} + \varepsilon_{2t}$. Where ε_{1t} and ε_{2t} are i.i.d. and normally distributed $(0,1)$.
2. Nonlinear with exponential component. $X_t = 1.4 - 0.5X_{t-1}e^{Y_{t-1}} + \varepsilon_{1t}$ and $Y_t = 0.4 + 0.23Y_{t-1} + \varepsilon_{2t}$; where ε_{1t} and ε_{2t} are i.i.d. normal $(0,1)$.
3. NLAR (Autoregressive Nonlinear). $X_t = 0.2|X_{t-1}|/(2 + |X_{t-1}|) + \varepsilon_{1t}$ and $Y_t = 0.7|Y_{t-1}|/(1 + |X_{t-1}|) + \varepsilon_{2t}$; where ε_{1t} and ε_{2t} are i.i.d. normal $(0,1)$.
4. Lorenz: $X_t = 1.96X_{t-1} - 0.8X_{t-1}Y_{t-1}$; $Y_t = 0.2Y_{t-1} + 0.8X_{t-1}^2$; with initial conditions X_1, Y_1 generated randomly. This is a discrete version of the Lorenz process as in [63].

Table 2 shows the results of the power experiments applying the SNC and the Granger noncausality test to 10,000 Monte Carlo simulations for the four models and for different sample sizes ($T = 50, 100, 500, 1000, \text{ and } 5000$).

Sample size	Model	Symbolic noncausality		Granger noncausality	
		$X \rightarrow Y$	$Y \rightarrow X$	$X \rightarrow Y$	$Y \rightarrow X$
$T = 50$	AR(1) (None)	0.40	0.45	5.66	5.25
$T = 100$		0.42	0.37	5.28	5.09
$T = 500$		0.41	0.27	5.47	5.14
$T = 1000$		0.39	0.42	5.18	5.14
$T = 5000$		0.34	0.46	5.30	5.09
$T = 50$	NLAR ($X \rightarrow Y$)	0.01	0.01	0.05	0.05
$T = 100$		0.73	0.34	4.82	4.74
$T = 500$		5.96	0.31	4.94	5.16
$T = 1000$		17.87	0.29	5.01	5.05
$T = 5000$		98.02	0.39	6.51	5.00
$T = 50$	Nonlinear exponential ($Y \rightarrow X$)	0.51	3.76	2.89	16.89
$T = 100$		0.28	11.85	2.78	13.36
$T = 500$		0.43	89.50	2.53	11.48
$T = 1000$		0.40	99.22	2.73	11.29
$T = 5000$		1.42	100.00	2.67	11.19
$T = 50$	Lorenz ($X \rightarrow Y, Y \rightarrow X$)	96.61	31.90	30.77	13.86
$T = 100$		99.99	90.49	28.52	12.64
$T = 500$		100.00	100.00	23.60	11.74
$T = 1000$		100.00	100.00	24.42	11.69
$T = 5000$		100.00	100.00	23.87	11.52

Table 2. Simulated power of the SNC and the Granger non causality statistic.

Following [60], a 60% acceptance or rejection of the null hypothesis is considered as a threshold. SNC and Granger noncausality correctly identifies noncausality in AR(1) process. **Table 2** suggests that SNC is more conservative in the rejection of causality with percentages less than 5%. The nonlinear model with an exponential component implies causality from Y to X . Note that SNC detects the causality when the sample size is 500 or larger. However, Granger test does not detect causality in any case. As asserted by [58] the NLAR process is very difficult to detect. Note that SNC is the only one detecting the causality when $T = 5000$. The Lorenz discrete map is also chaotic, and it is detected by SNC starting from $T = 100$. However, note that Granger test never detects the causality. In particular, is highlighted that Granger test is not able to detect the model with an exponential component, the NLAR model and the chaotic Lorenz map.

Finally, we compare both tests with real data from US. In particular, we consider two well-known relationships in economics: the Phillips curve [64] about the relation between unemployment and inflation rates, the Okun's law [65] establishing a relation between unemployment and economic rate. We take annual data for the US unemployment rate, inflation rate,

and economic growth for the period 1948–2016 representing a total of 69 observations. **Table 3** shows the results of the Granger noncausality test and the symbolic test considering a partition of two symbols.

The results are similar for both tests. On one hand, Granger and symbolic tests detect causality from inflation to unemployment in the Phillips curve. On the other hand, the two tests detect causality running from economic growth to unemployment in the Okun’s law. The economic theory suggests that inflation increases unemployment while economic growth reduces it. Note that STSA allows thinking about causality in a more general way, whereas Granger noncausality needs to think of continuous measured variables, this should not be a problem for STSA. Let us consider the following example; we now can test the hypothesis of causality from economic growth (G) and inflation (P) to unemployment (U). The main problem is that we have to test causality from a two-dimensional variable to a one dimensional. Symbolization permits to transform the two-dimensional problem in one dimensional and then to apply the symbolic test as explained. We can follow a similar approach as in [66] where STSA is applied

Null hypothesis	Granger	SNC(2 symbols)
Phillips curve		
Unemployment does not cause inflation	0.04	1.53
Inflation does not cause unemployment	16.90*	9.41*
Okun’s law		
Unemployment does not cause economic growth	3.37	2.94
Economic growth does not cause unemployment	61.01*	9.65*

* Indicates rejection of the null hypothesis at the 5% level significance.

Table 3. SNC and the Granger non causality for the Phillips Curve and Okun’s Law in US.

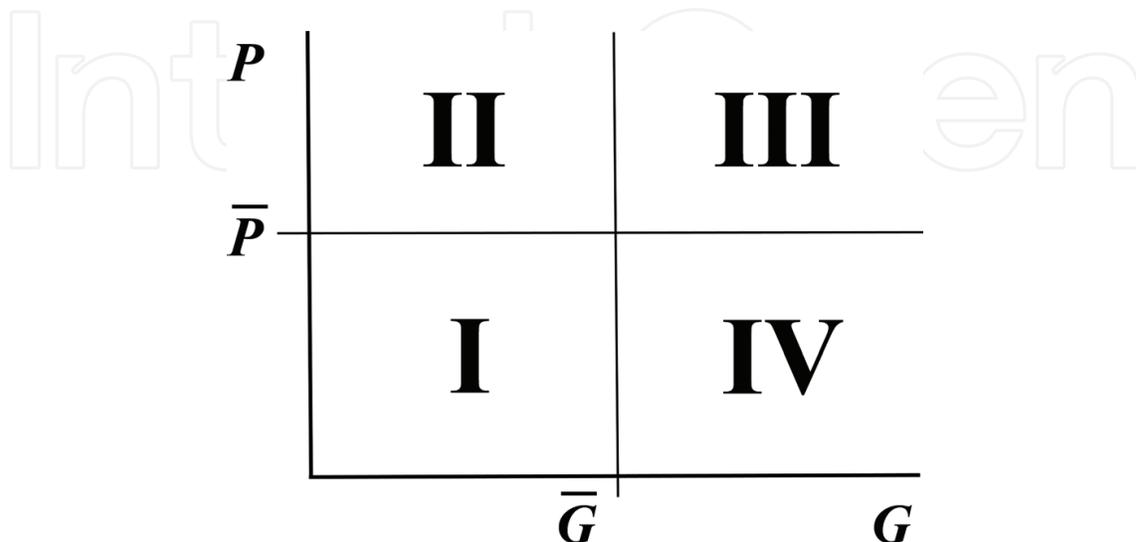


Figure 2. Two-dimensional variable (economic growth and inflation) is transformed into a four symbol variable.

to dynamic regimes. **Figure 2** shows the transformation of the variable (G, P) in a symbolic variable with an alphabet of four symbols (I: low economic growth and low inflation, II: low economic growth and high inflation, III: high economic growth and high inflation, IV: high economic growth and low inflation) considering as partition the mean of each variable. Note that now the application of symbolic causality is easy, the hypothesis that the economic regime (G, P) does not cause unemployment is rejected since the SNC is 31.76 and Chi-2 with 15 degree of freedom (4^2-1) at 95% is 25.00. The opposite hypothesis is not rejected because the SNC is 24.71. It is not possible to test this type of causality with the traditional Granger noncausality test.

8. Conclusion

STSA is a powerful tool being applied to many scientific fields. There are recent applications in robotic, biology, medicine, communication, and engineering. However, applications in social sciences are very recent. The main difficult is the few historical data produced by the social processes. Social sciences are used to applied statistical tests for proving their hypothesis. However, there is much work to do in developing statistical tests based on STSA to be applied in social sciences. There are some very recent efforts applied to economics and finance using STSA. In particular, we present a symbolic independence test, which seems to be powerful in detecting nonlinearities compared with well-known BDS and runs test. The symbolic test is better detecting models such as the chaotic Anosov and Logistic or some stochastic models such as NLMA or NLSIGN. A second symbolic test about causality detects complex processes such as NLAR, nonlinear exponential, or the Lorenz chaotic process when the traditional Granger noncausality cannot. The symbolic causality also enables causality to be tested in a more general perspective. The application of test from a two-dimensional economic variable to a one-dimensional economic variable is a clear example of the potential of STSA in economics and social sciences in general.

One future research line could be to develop a powerful nonlinear test for multidimensional variables. As it was explained, STSA permits to transform a multidimensional time series in a one-dimensional time series simplifying the analysis. This could have important applications in relationships involving vector functions. A more general line of research is to find methodologies to define the optimal partition. As mentioned before, equiprobable partition is generally applied but to find the right partition is still a theoretical and practical weakness in STSA.

Author details

Wiston Adrián Risso

Address all correspondence to: arisso@iecon.cce.edu.uy

Institute of Economics (IECON), University of the Republic, Montevideo, Uruguay

References

- [1] Box G, Jenkins G. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day ed; 1970
- [2] Sims C. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*. 1980;**48**(1):1-48
- [3] Engle R. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*. 1982;**50**(4): 987-1007
- [4] Engle R, Granger C. Co-integration and error correction: Representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*. 1987;**55**(2):251-276
- [5] Granger C. Investigating causal relations by econometrics models and cross-spectral methods. *Econometrica*. 1969;**37**(3):424-438
- [6] Box G, Jenkins G, Reinsel G, Ljung G. Time Series Analysis: Forecasting and Control. 5th ed. John Wiley & Sons: New Jersey, US; 2015
- [7] Granger C, Newbold P. Forecasting Economic Time Series. Academic Press: New York, US; 2014
- [8] Hamilton J. Time Series Analysis. Princeton: Princeton University; 1994
- [9] Harris R, Sollis R. Applied Time Series Modelling and Forecasting. Wiley: Chichester, UK; 2003
- [10] Franses P. Time Series Models for Business and Economic Forecasting. Cambridge University Press: Cambridge, UK; 1998
- [11] Chatfield C. Time-Series Forecasting. CRC Press: Florida, US; 2000
- [12] Lütkepohl H. New Introduction to Multiple Time Series Analysis. Springer Science & Business Media: Berlin, Germany; 2005
- [13] Clements M, Hendry D. Forecasting Economic Time Series. Cambridge University Press: Cambridge, UK; 1998
- [14] Harris R. Using Cointegration Analysis in Econometric Modelling (Vol. 82). London: Prentice Hall; 1995
- [15] Banerjee A, Dolado JJ, Galbraith JW, Hendry D. Co-integration, Error Correction, and The Econometric Analysis of Non-Stationary Data. OUP Catalogue: Oxford, UK; 1993
- [16] Juselius K. The Cointegrated VAR Model: Methodology and Applications. Oxford University Press: Oxford, UK; 2006
- [17] Johansen S. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press: Oxford, UK; 1995

- [18] Tsay R. *Analysis of Financial Time Series* (Vol. 543). John Wiley & Sons: New York, US; 2005
- [19] Terasvirta T, Tjostheim D, Granger CW. *Modelling Nonlinear Economic Time Series*. OUP Catalogue: Oxford, UK; 2010
- [20] Kantz H, Schreiber T. *Nonlinear Time Series Analysis* (Vol. 7). Cambridge University Press: Cambridge, UK; 2004
- [21] Williams S. Symbolic dynamics and its applications. *Proceeding of Symposia in Applied Mathematics*. 2004;**60**:150
- [22] Piccardi C. On the control of chaotic systems via symbolic time series analysis. *Chaos*. 2004;**14**(4):1026-1034
- [23] Daw C, Finney C, Tracy E. A review of symbolic analysis of experimental data. *Review of Scientific Instruments*. 2003;**74**(2):915-930
- [24] Daw CS, Kennel MB, Finney CEA, Connolly F. Observing and modeling nonlinear dynamics in an internal combustion engine. *Physical Review E*. 1998;**57**(3):2811
- [25] Daw C, Green J, Wagner R, Finney C, Connolly F. Synchronization of combustion variations in a multicylinder spark ignition engine. *Proceedings of the Combustion Institute*. 2000;**28**(1):1249-1255
- [26] Daw C, Finney C, Kennel M. Symbolic approach for measuring temporal "irreversibility". *Physical Review E*. 2000;**62**(2):1912
- [27] Goldstein R. A technique for the measurement of the power spectra of very weak signals. *IRE Transactions on Space Electronics and Telemetry*. 1962;**2**:170-173
- [28] Schwarz U, Benz A, Kurths J, Witt A. Analysis of solar spike events by means of symbolic dynamics methods. *Astronomy and Astrophysics*. 1993;**277**:215
- [29] Hively L, Gailey P, Protopopescu V. Detecting dynamical change in nonlinear time series. *Physics Letters A*. 1999;**258**(2):103-114
- [30] Hively L, Protopopescu V, Gailey P. Timely detection of dynamical change in scalp EEG signals. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2000;**10**(4):864-875
- [31] Makki M, Samali B, Li J. Damage localization based on symbolic time series analysis. *Structural Control and Health Monitoring*. 2015;**22**(2):374-393
- [32] Qumar A, Aziz W, Saeed S, Ahmed I, Hussain L. Comparative study of multiscale entropy analysis and symbolic time series analysis when applied to human gait dynamics. In: 2013 International Conference on Open Source Systems and Technologies (ICOSST). IEEE: Lahore, Pakistan; December, 2013; 2013.
- [33] Kim J, Park J, Seo J, Lee W, Kim H, Noh J, Yum M. Decreased entropy of symbolic heart rate dynamics during daily activity as a predictor of positive head-up tilt test in patients with alleged neurocardiogenic syncope. *Physics in Medicine and Biology*. 2003;**45**(11):3403

- [34] Cammarota C, Rogora E. Time reversal, symbolic series and irreversibility of human heartbeat. *Chaos, Solitons & Fractals*. 2007;**32**(5):1649-1654
- [35] Rao A, Jain R. Computerized flow field analysis: Oriented texture fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1992;**14**(7):693-709
- [36] Edwards R, Siegelmann H, Aziza K, Glass L. Symbolic dynamics and computation in model gene networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2001;**11**(1):160-169
- [37] Godelle J, Letellier C. Symbolic sequence statistical analysis for free liquid jets. *Physical Review E*. 2000;**62**(6):7973
- [38] Sarkar S, Mukherjee K, Sarkar S, Ray A. Symbolic transient time-series analysis for fault detection in aircraft gas turbine engines. In: *American Control Conference (ACC)*; June 2002; 2002. pp. 5132–5137.
- [39] Ramanan V, Chakravarthy SR, Sarkar S, Ray A Investigation of combustion instability in a swirl-stabilized combustor using symbolic time series analysis. In: *American Society of Mechanical Engineers*, editor. *ASME 2014 Gas Turbine India Conference*; December 2014; 2014. p. V001T03A012-V001T03A012.
- [40] Hsu A, Marshall A, Ricca T. Clipped representations of fourier-transform ion-cyclotron resonance mass spectra. *Analytica Chimica Acta*. 1985;**178**:27-41
- [41] Baptista M, Rosa E, Grebogi C. Communication through chaotic modeling of languages. *Physical Review E*. 2000;**61**(4):3590
- [42] Dolnik M, Bollt E. Communication with chemical chaos in the presence of noise. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 1998;**8**(3):702-710
- [43] Seto Y, Takahashi N, Jha D, Virani N, Ray A Data-driven robot gait modeling via symbolic time series analysis. In: *American Control Conference (ACC)*; July 2016; 2016. p. 3904–3909.
- [44] Gullapalli A, Carley K. Extracting ordinal temporal trail clusters in networks using symbolic time-series analysis. *Social Network Analysis and Mining*. 2013;**3**(4):1179-1194
- [45] Das G, Lin KI, Mannila H, Renganathan G, Smyth P. Rule discovery from time series. *KKD*. 1998;**98**(1):16-22
- [46] Brida J, Risso W. Dynamics and structure of the main italian companies. *International Journal of Modern Physics C*. 2007;**18**(11):1783-1793
- [47] Brida J, Risso W. Multidimensional minimal spanning tree: The dow jones case. *Physica A: Statistical Mechanics and its Applications*. 2008;**387**(21):5205-5210
- [48] Brida J, Risso W. Dynamic and structure of the Italian Stock Market based on returns and volume trading. *Economics Bulletin*. 2009;**29**(3):2417-2423
- [49] Brida J, Gomez D, Risso W. Symbolic hierarchical analysis in currency markets. An application to contagion in currency crises. *Expert Systems with Applications*. 2009;**36**(4):7721-7728

- [50] Brida J, Esteban L, Risso W, Such S. The international hotel industry in Spain: Its hierarchical structure. *Tourism Management*. 2010;**31**(1):57-73
- [51] Risso W. The informational efficiency and the financial crashes. *Research in International Business and Finance*. 2008;**22**(3):396-408
- [52] Risso W. The informational efficiency: The emerging markets versus the developed markets. *Applied Economics Letters*. 2009;**16**(5):485-487
- [53] Clausius R. The nature of the motion we call heat. In: Brush SG, editor. *Kinetic Theory. (Selected Readings in Physics)*. Vol. Vol. 2. Pergamon Press: Oxford, UK; 1965. p. 111-134
- [54] Shannon C. A mathematical theory of communication. *Bell System Technical Journal*. 1948;**27**:379-423
- [55] Khinchin A. *Mathematical Foundations of Information Theory*. Courier Dover Publications: New York, US; 1957
- [56] Wald A, Wolfowitz J. An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*. 1943;**14**(4):378-388
- [57] Brock W, Dechert W, LeBaron B, Scheinkman J. A test for independence based on the correlation dimension. *Econometric Reviews*. 1996;**15**:197-235
- [58] Risso W. An independence test based on symbolic time series. *International Journal of Statistical Mechanics*. 2014;**2014**:809383
- [59] Mathai A, Provost S. *Quadratic Forms in Random Variables: Theory and Applications*. Marcel Dekker, Inc.; 1992
- [60] Risso W. A first approach on testing non-causality with symbolic time series. *Economic Computation and Economic Cybernetics Studies and Research*. 2015;**49**(3):123-142
- [61] Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*. 2007;**441**(1): 1-46
- [62] Cover T, Thomas J. *Elements of Information Theory*, Wiley. 2nd ed 2006
- [63] Stork, M., Hrusak, J., Mayer, D.. Discrete-time chaotic systems impulsive synchronization and data transmission. In: *Proc. 13th WSEAS Int. Conf. System*; 2009.
- [64] Phillips A. The relationship between unemployment and the rate of change of money wages in the United Kingdom 1861–1957. *Economica*. 1958;**25**(100):283-299. DOI: 10.1111/j.1468-0335.1958.tb00003.x
- [65] Okun A. Potential GNP, its measurement and significance. *Proceedings of the Business and Economics Statistics Section of the American Statistical Association*. 1962:98-104
- [66] Brida J, Punzo L. Symbolic time series analysis and dynamic regimes. *Structural Change and Economic Dynamics*. 2003;**14**(2):159-183

