



# Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency

Luciano Zunino<sup>a,b,c,\*</sup>, Massimiliano Zanin<sup>d</sup>, Benjamin M. Tabak<sup>e,f</sup>, Darío G. Pérez<sup>g</sup>, Osvaldo A. Rosso<sup>h,i</sup>

<sup>a</sup> Instituto de Física Interdisciplinaria y Sistemas Complejos (IFISC) CSIC-UIB, Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

<sup>b</sup> Centro de Investigaciones Ópticas (CONICET La Plata - CIC), C.C. 3, 1897 Gonnet, Argentina

<sup>c</sup> Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina

<sup>d</sup> Universidad Autónoma de Madrid, 28049 Madrid, Spain

<sup>e</sup> Banco Central do Brasil - SBS Quadra 3, Bloco B, 9 andar, DF 70074-900, Brazil

<sup>f</sup> Universidade Católica de Brasília, Brasília, DF, Brazil

<sup>g</sup> Instituto de Física, Pontificia Universidad Católica de Valparaíso (PUCV), 23-40025 Valparaíso, Chile

<sup>h</sup> Instituto de Ciências Exatas (Física), Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 - Campus Pampulha, 31270-901 Belo Horizonte - MG, Brazil

<sup>i</sup> Chaos & Biology Group, Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón II, Ciudad Universitaria. 1428 Ciudad de Buenos Aires, Argentina

## ARTICLE INFO

### Article history:

Received 28 August 2009

Received in revised form 14 December 2009

Available online 15 January 2010

### Keywords:

Complexity-entropy causality plane

Bandt and Pompe method

Stock market inefficiency

## ABSTRACT

The complexity-entropy causality plane has been recently introduced as a powerful tool for discriminating Gaussian from non-Gaussian process and different degrees of correlations [O.A. Rosso, H.A. Larrondo, M.T. Martín, A. Plastino, M.A. Fuentes, Distinguishing noise from chaos, *Phys. Rev. Lett.* 99 (2007) 154102]. We propose to use this representation space to distinguish the stage of stock market development. Our empirical results demonstrate that this statistical physics approach is useful, allowing a more refined classification of stock market dynamics.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

The understanding and analysis of economical time series, especially the evolution of price sequences, has been attracting the attention of mathematicians and physicists for many years. The French mathematician Louis Bachelier, almost a century ago, modeled the market prices by what today is known as a random walk [1]. This cornerstone of the financial modelling is based on the assumptions that the increments of the prices are independent and obey a Gaussian distribution. Later, in 1970, Eugene Fama introduced the celebrated Efficient Market Hypothesis (EMH) [2]. According to the weak-form version of this hypothesis,<sup>1</sup> at any given time, the price of an asset fully reflects all available historical information, but it is widely known that this is only a first approximation. Deviations from this model, violating either the independence or Gaussian assumptions, have been found in many empirical studies since the revolutionary paper of Benoit Mandelbrot [3]. The existence of autocorrelation between distant observations breaks the market efficiency because past prices can help to

\* Corresponding author at: Instituto de Física Interdisciplinaria y Sistemas Complejos (IFISC) CSIC-UIB, Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain.

E-mail addresses: [lucianoz@ciop.unlp.edu.ar](mailto:lucianoz@ciop.unlp.edu.ar) (L. Zunino), [massimiliano.zanin@hotmail.com](mailto:massimiliano.zanin@hotmail.com) (M. Zanin), [benjamin.tabak@bcb.gov.br](mailto:benjamin.tabak@bcb.gov.br) (B.M. Tabak), [dario.perez@ucv.cl](mailto:dario.perez@ucv.cl) (D.G. Pérez), [oarosso@fibertel.com.ar](mailto:oarosso@fibertel.com.ar) (O.A. Rosso).

<sup>1</sup> The semi-strong and strong form of EMH take into account other informations and not only the historical prices. More precisely, the semi-strong approach considers that prices reflect all public information and the strong version states that prices reflect any information.

predict future prices; i.e., correlated markets allow for arbitrage opportunities. This question motivates the research on the subject, especially by portfolio managers and analysts, trying to generate higher returns. Bęben and Orłowski [4] and Di Matteo et al. [5,6] have found that emergent markets have greater correlation than developed markets, suggesting more predictability. Thus, emergent economies seem to be less efficient than developed ones. More recently, a negative correlation between stock market efficiency and predictability was experimentally assessed [7,8].

Nowadays it is clear that residual inefficiencies are always present in actual stock markets and that the concept of market efficiency is only an idealization [9,10]. Concepts like economic management, political stability and risk profile distinguish emergent and developed stock markets. It was also shown that liquidity and market capitalization play an important role in understanding stock market inefficiency [11]. However, these properties are very difficult to quantify in a precise way [9].

Several authors have proposed the Hurst parameter as a measure of the stock market inefficiency [6–8,11–18]. More precisely, they argue that this scaling exponent measures the long-range dependence exhibited by the stock market indices under analysis. However, Bassler et al. [19] have recently shown that the estimation of the Hurst parameter alone cannot be used to determine either the existence of long-term memory or the efficiency of markets. Moreover, they found that Hurst parameters  $H \neq 1/2$  are perfectly consistent with Markov processes and the EMH. Markov processes, by construction, are memoryless and, in this case,  $H \neq 1/2$  basically means that the associated process has nonstationary increments. Thus, they conclude that for an appropriately evidence for autocorrelations in data it is necessary to show that the market price increments are stationary besides an empirical measurement or theoretical prediction of the Hurst parameter. Other approaches have been recently introduced to rank stock markets in order to distinguish between emergent and developed economies [9,10,20–27]. In particular, Complex Systems and Nonlinear Dynamics theories have been used to shed new light on this class of problems, thanks to their abilities to detect hidden structures inside noisy information, even when the considered system exhibits a behavior at the edge of chaos [28].

In this paper we propose to test the market efficiency of financial time series by employing a recently introduced statistical tool; the *complexity-entropy causality plane*. It was shown that this plane allows to distinguish (i) Gaussian from non-Gaussian process and (ii) different degrees of correlations [29]. We empirically show that it is a powerful physical tool for discriminating market dynamics.

## 2. Complexity-entropy causality plane

The concept of entropy can be of great help when analyzing stock market data since it captures the uncertainty and disorder of the time series without imposing any limitations on the theoretical probability distribution [30,31]. If prices were a pure random walk, the variations would be a completely uncorrelated string of numbers. We would say that such string of data is completely disordered and its entropy is maximized. On the other hand, if the price variations are somewhat correlated, then the entropy does not attain its maximal value. Therefore, the negative entropy (i.e., the entropy of price variations with reference to its maximal value) can be taken as a measure of predictability and, consequently, of market inefficiency [32]. We should remark that the Shannon, Renyi, Tsallis [31,33,34], Approximate [21,35] and transfer [36] entropies, and the Shannon entropy local approach [37] have been proposed to quantify different aspects of financial time series.

The information content of a system is typically evaluated from a probability distribution function,  $P$ , describing the distribution of some measurable or observable property. An information measure can primarily be viewed as a quantity that characterizes this given probability distribution. In this work the Shannon entropy is used as a first natural approach. Given any arbitrary probability distribution  $P = \{p_i : i = 1, \dots, N\}$ , the widely known Shannon's logarithmic information measure defined by

$$S[P] = - \sum_{i=1}^N p_i \ln p_i \quad (1)$$

is regarded as the measure of the uncertainty associated to the physical processes described by  $P$ . If  $S[P] = 0$  we are in position to predict with certainty which of the possible outcomes  $i$  whose probabilities are given by the  $p_i$  will actually take place. Our knowledge of the underlying process described by the probability distribution is in this instance maximal. On the other hand, our ignorance is maximal for a uniform distribution.

Nevertheless, Lamberti et al. [38] introduced a statistical complexity measure (SCM) able to detect essential details of the dynamics and differentiate different degrees of periodicity and chaos. It provides important additional information regarding the peculiarities of the underlying probability distribution, not already detected by the entropy. This statistical complexity measure is defined through the product

$$C_S[P] = \mathcal{Q}_J[P, P_e] \mathcal{H}_S[P] \quad (2)$$

of the normalized Shannon entropy

$$\mathcal{H}_S[P] = S[P]/S_{\max} \quad (3)$$

– with  $S_{\max} = S[P_e] = \ln N$ , ( $0 \leq \mathcal{H}_S \leq 1$ ) and  $P_e = \{1/N, \dots, 1/N\}$  the uniform distribution – and the disequilibrium  $\mathcal{Q}_J$  defined in terms of the extensive (in the thermodynamical sense) Jensen–Shannon divergence [38]. That is,

$$\mathcal{Q}_J[P, P_e] = \mathcal{Q}_0 \mathcal{J}[P, P_e] \quad (4)$$

with  $\mathcal{J}[P, P_e] = \{S[(P+P_e)/2] - S[P]/2 - S[P_e]/2\}$  the above-mentioned Jensen–Shannon divergence and  $\mathcal{Q}_0$  a normalization constant, equal to the inverse of maximum possible value of  $\mathcal{J}[P, P_e]$ ; i.e., the value obtained when one of the component of  $P$ , say  $p_m$ , equals unity and the remaining  $p_i$  vanish. The complexity measure constructed in this way has the intensive property found in many thermodynamic quantities.

The system architecture manifests through the SCM, since it is different from zero only when there exist privileged, more likely, states among the accessible ones [39]. Not only does it quantify randomness, but also the presence of correlation among components of the dynamical system [38,40]. The opposite extremes, perfect order and maximal randomness (a periodic sequence and a fair coin toss, for example) possess no structure; these systems are too simple and have zero statistical complexity. In between these two special instances, a wide range of possible degrees of physical structure exist, degrees that should be reflected in the features of the underlying probability distribution. We insist on the fact that the above SCM is not a trivial function of the entropy because it depends on two different probabilities distributions: the one associated to the system under analysis,  $P$ , and the uniform distribution,  $P_e$ . Furthermore, it was shown that for a given  $\mathcal{H}_S$  value, there exists a range of possible SCM values between a minimum  $C_{\min}$  and a maximum  $C_{\max}$  [41]. It was recently shown that measures of statistical or structural complexity are necessary because they capture the property of organization [42]. This kind of information is not discriminated by randomness measures.

Finally, in order to study the time evolution of the SCM, a diagram of  $C_S$  versus  $\mathcal{H}_S$  can be used. As is widely known, the second law of thermodynamics states that entropy grows monotonically with time. Thus,  $\mathcal{H}_S$  can be regarded as an arrow of time. This diagnostic tool measures the system’s intrinsic computation in a parameter-free way [42]. It has been successfully used to study changes in a system dynamics originated by modifications of some characteristic parameters [38,39,43–46].

### 2.1. Bandt & Pompe symbolization method

In order to evaluate the two above-mentioned quantifiers an associated probability distribution should be constructed beforehand. The adequate way of picking up the probability distribution associated to a time series is an open problem. Rosso et al. [29] have recently shown that improvements in the performance of information quantifiers, like entropy and statistical complexity measures, can be expected if the underlying probability distribution is obtained taking into account the time causality of the system’s dynamics. In particular, it was found that these information measures allow to distinguish between chaotic and stochastic dynamics when causal information is incorporated into the scheme to generate the associated probability distribution. Bandt & Pompe [47] introduced a successful method to evaluate the probability distribution considering this time causality. The Bandt & Pompe method (BPM) is based on the details of the phase space reconstruction procedure and it is the only one among those in popular use that takes into account the temporal structure of the time series generated by the physical process under study. Then, important details concerning the ordinal structure of the time series can be revealed [48–51].

Given a time series  $\{x_t : t = 1, \dots, M\}$ , an embedding dimension  $D > 1$ , and a time delay  $\tau$ , consider the ordinal patterns of order  $D$  [47,48,52] generated by

$$s \mapsto (x_{s-(D-1)\tau}, x_{s-(D-2)\tau}, \dots, x_{s-\tau}, x_s). \tag{5}$$

To each time  $s$  we assign a  $D$ -dimensional vector that results from the evaluation of the time series at times  $s, s - \tau, \dots, s - (D - 1)\tau$ . Clearly, the greater the  $D$  value, the more information about the past is incorporated into the ensuing vectors. By the ordinal pattern of order  $D$  related to the time  $s$  we mean the permutation  $\pi = (r_0, r_1, \dots, r_{D-1})$  of  $(0, 1, \dots, D - 1)$  defined by

$$x_{s-r_{D-1}\tau} \leq x_{s-r_{D-2}\tau} \leq \dots \leq x_{s-r_1\tau} \leq x_{s-r_0\tau}. \tag{6}$$

In this way the vector defined by Eq. (5) is converted into a unique symbol  $\pi$ . Let us make a numerical example; we start with the time series  $\{1, 3, 5, 4, 2, 5, \dots\}$ , and we set the embedding dimension  $D = 4$  and the time delay  $\tau = 1$ . In this case the state space is divided into  $4!$  partitions and 24 mutually exclusive permutation symbols are considered. The first 4-dimensional vector will be  $(1, 3, 5, 4)$ . Then, the ordinal pattern which allows us to fulfill Eq. (6) will be  $(1, 0, 2, 3)$ . The second vector will be  $(3, 5, 4, 2)$ , and  $(2, 1, 3, 0)$  will be its associated permutation, and so on. In order to get a unique result we consider that  $r_i < r_{i-1}$  if  $x_{s-r_i\tau} = x_{s-r_{i-1}\tau}$ . This is justified if the values of  $x_t$  have a continuous distribution so that equal values are very unusual. Thus, for all the  $D!$  possible permutations  $\pi_i$  of order  $D$ , their associated relative frequencies can be naturally computed by

$$p(\pi_i) = \frac{\#\{s | 1 + (D - 1)\tau \leq s \leq M \text{ and } s \text{ has ordinal pattern } \pi_i\}}{M - (D - 1)\tau}, \tag{7}$$

where  $\#$  is the cardinality of the set—roughly speaking, the number of elements in it. Thus, an ordinal pattern probability distribution  $P = \{p(\pi_i), i = 1, \dots, D!\}$  is obtained from the time series. To determine  $p(\pi_i)$  exactly an infinite time series should be considered, taking  $M \rightarrow \infty$  in the above formula. This limit exists with probability 1 when the underlying stochastic process fulfills a very weak stationarity condition: for  $k \leq D$ , the probability for  $x_t < x_{t+k}$  should not depend on  $t$  [47]. The probability distribution  $P$  is obtained once we fix the embedding dimension  $D$  and the time delay  $\tau$ . The former parameter plays an important role for the evaluation of the appropriate probability distribution, since  $D$  determines

**Table 1**

World stock indices. Emergent and developed stock markets following the MSCI methodology are denoted as E and D, respectively.

Country	Number of observations	Bloomberg code	Index
1. Argentina (E)	3117	MERVAL	Argentina Merval Index
2. Australia (D)	3178	ASX	Australian Securities Exchange
3. Austria (D)	3109	ATX	Austrian Traded Index
4. Belgium (D)	3179	BEL20	BEL20 Index
5. Brazil (E)	3109	IBOV	Brazil Bovespa Stock Index
6. Canada (D)	3163	SPTSX	S&P/TSX Composite Index
7. Chile (E)	3133	IPSA	Chile Stock Market Select Index
8. China (E)	3034	SHCOMP	Shanghai Stock Exchange Composite Index
9. Denmark (D)	3147	KFX	OMX Copenhagen 20 Index
10. France (D)	3179	CAC	CAC 40 Index
11. Germany (D)	3176	DAX	DAX Index
12. Greece (D)	3141	ASE	ASE General Index
13. Hong Kong (D)	3099	HSI	Hang Seng Index
14. India (E)	3098	SENSEX	Bombay Stock Exchange Sensitive Index
15. Indonesia (E)	3073	JCI	Jakarta Composite Index
16. Ireland (D)	3153	ISEQ	Irish Overall Index
17. Italy (D)	3180	MIB30	Milan MIB30 Index
18. Japan (D)	3093	TPX	Tokio Stock Price Index
19. Korea (E)	3272	KOSPI	Korea Composite Index
20. Malaysia (E)	3093	KLCI	Kuala Lumpur Composite Index
21. Mexico (E)	3156	MEXBOL	Mexico Bolsa Index
22. Philippines (E)	3124	PCOMP	Philippines Composite Index
23. Singapore (D)	3149	STI	Straits Times Index
24. Spain (D)	3156	IBEX	IBEX 35 Index
25. Sweden (D)	3149	OMX	OMX Stockholm 30 Index
26. Switzerland (D)	3159	SMI	Swiss Market Index
27. Taiwan (E)	3282	TWSE	Taiwan Taiex Index
28. Thailand (E)	3080	SET	Stock Exchange of Thailand Index
29. Turkey (E)	3120	XU100	Istanbul Stock Exchange National 100 Index
30. United Kingdom (D)	3172	UKX	FTSE 100 Index
31. US (D)	3161	SPX	S&P 500 Index
32. Venezuela (E)	3023	IBVC	Venezuela Stock MKT Index

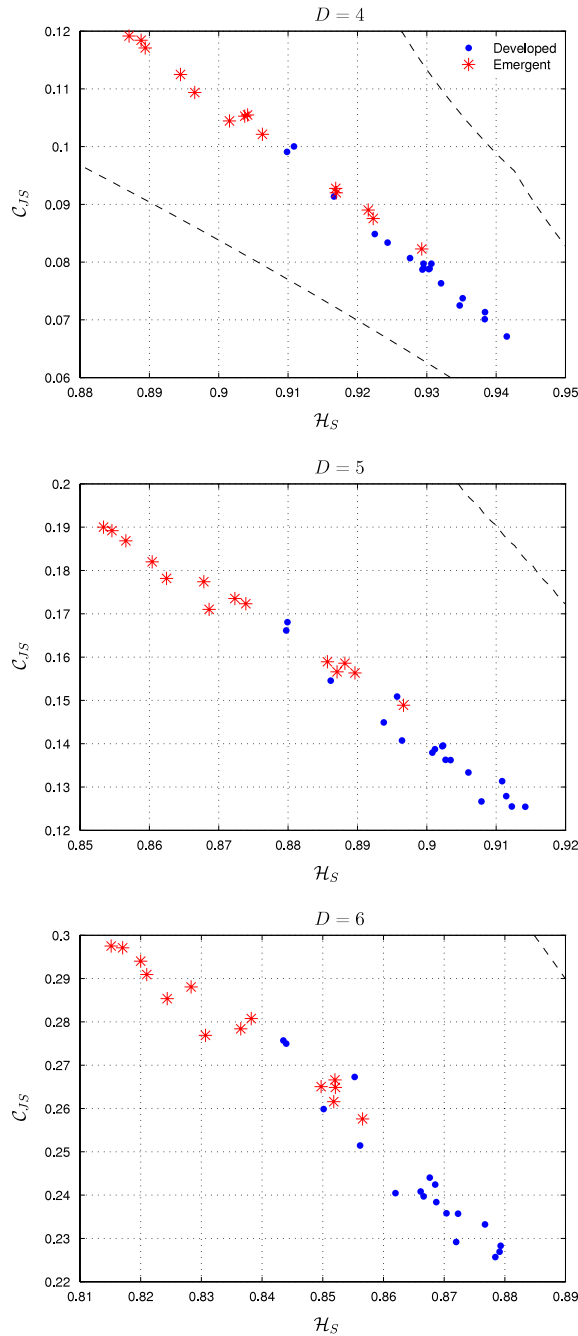
the number of accessible states,  $D!$ . Moreover, it was established [53] that the length  $M$  of the time series must satisfy the condition  $M \gg D!$  in order to work with a reliable statistics. In particular, Bandt & Pompe suggest for practical purposes to work in the range  $3 \leq D \leq 7$ . They specifically considered a time delay  $\tau = 1$  in their cornerstone paper [47]. Nevertheless, it is clear that other values could give additional information. Very recently, Matilla-García and Ruiz Marín have introduced a helpful method to detect lags in nonlinear time series by considering different time delays [54].

In this paper we evaluate the normalized Shannon entropy,  $\mathcal{H}_S$  Eq. (3), and the intensive SCM,  $\mathcal{C}_{JS}$  (Eq. (2)), using the permutation probability distribution,  $P = \{p(\pi_i), i = 1, \dots, D!\}$ , introduced in this section. The diagram of  $\mathcal{C}_{JS}$  versus  $\mathcal{H}_S$  by employing this probability distribution is the complexity-entropy causality plane defined in Ref. [29]. The Shannon entropy defined through the permutation probability distribution is usually known as permutation entropy. It has been used in several papers for different purposes [27,55–58]. In particular, Rosso and Masoller [59,60] have recently shown that both Information Theory quantifiers,  $\mathcal{H}_S$  and  $\mathcal{C}_{JS}$ , evaluated with the BPM's probability distribution, are suitable tools for detecting and quantifying noise-induced temporal correlations in stochastic resonance phenomena.

### 3. Empirical results

In this paper we analyze the prices of 32 equity indices for different countries. All data were collected from the Bloomberg database ([www.bloomberg.com](http://www.bloomberg.com)). The codes and names of these indices are presented in Table 1. We employ daily prices beginning on January 2, 1995 and ending in July 23, 2007. We have on average 3138 observations. Time counting was performed over trading days, skipping weekends and holidays. Exact numbers of observations for each country are detailed in the second column of Table 1. The choice of the beginning of the series is due to availability of data. Besides, in order to make comparisons, all country indices were studied for the same time period. Eighteen developed and fourteen emergent stock markets are considered. This classification is obtained following the Morgan Stanley Capital Index (MSCI) methodology to define developed and emergent stock markets ([www.msicibarra.com](http://www.msicibarra.com)). This is particularly useful as it allows a direct comparison of different classes of countries.

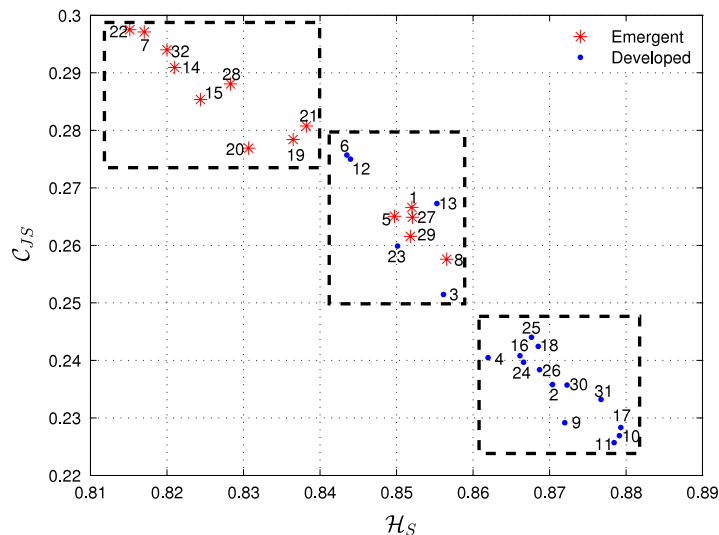
In this work we have analyzed daily prices; clearly, these time series are non-stationary. However the quantifiers used can be applied to non-stationary process. They have been used to characterize the fractional Brownian motion, a widely known non-stationary process—see Refs. [29,58,61] for further details. Moreover, we have previously shown that in the particular case of the normalized permutation entropy and in order to discriminate time series, better results are obtained for prices than log-returns [27]. However, this issue deserves special attention and will be the focus of a next study.



**Fig. 1.** Localization of the developed and emergent stock markets, according to the MSCI criterion, in the complexity-entropy causality plane with embedding dimensions  $D = 4$  (upper plot),  $D = 5$  (central plot) and  $D = 6$  (lower plot), and time delay  $\tau = 1$ . We also display the maximum and minimum possible values of the intensive statistical complexity measure (segmented lines). For further details about the range of possible SCM values see Ref. [41].

In Fig. 1 we have plotted the localization of the different countries in the complexity-entropy causality plane with embedding dimensions  $D = 4$ ,  $D = 5$  and  $D = 6$ , and time delay  $\tau = 1$ . Observe that developed countries have higher entropy and lower complexity than emergent ones. Therefore the former are closer to the ideal point ( $\mathcal{H}_S = 1$ ,  $\mathcal{C}_{JS} = 0$ ) and this conclusion is independent of the embedding dimension value. It should also be stressed that there is a clear dependence of  $\mathcal{C}_{JS}$  on  $\mathcal{H}_S$  and this relationship is independent of the particular embedding dimension considered, as one can easily conclude from Fig. 1. Comparing with Fig. 1 of Ref. [29] this behavior allows us to suggest that the prices can be modelled as a fractional Brownian motion.

We have 3138 observations on average. Thus, according to the condition introduced at the end of the previous section, the highest possible embedding dimension is 6 ( $3138 \gg 6! = 720$ ). This is the optimal value for taking into account



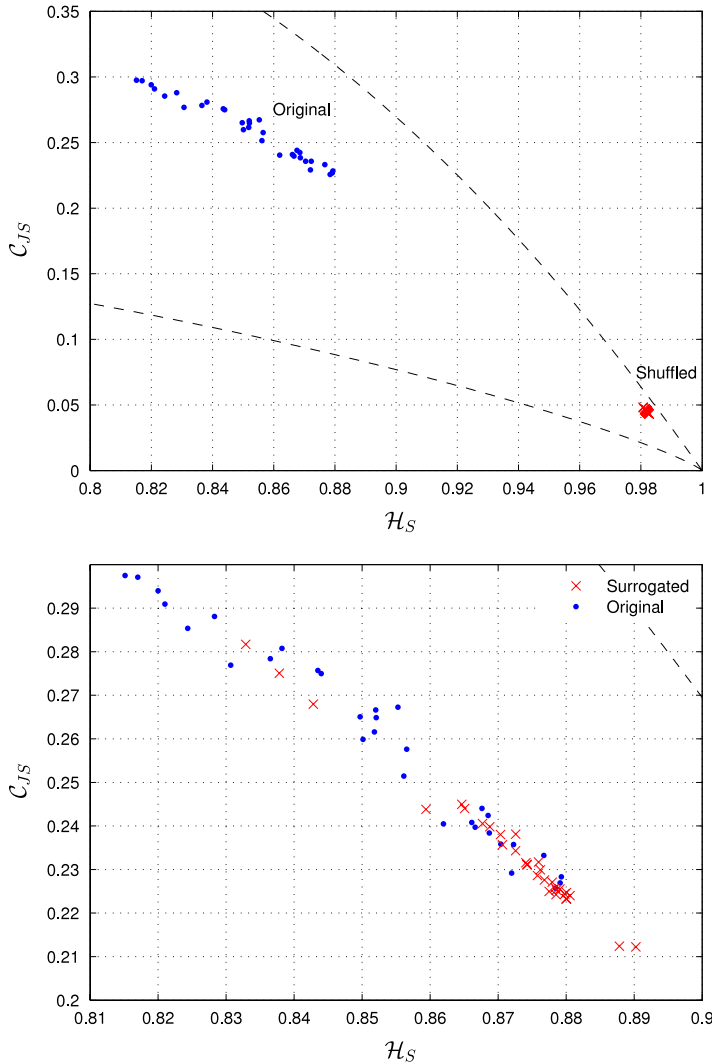
**Fig. 2.** Localization of the different stock markets in the complexity-entropy causality plane with embedding dimension  $D = 6$  and time delay  $\tau = 1$ . Numbers indicate stock markets listed in Table 1. The upper, central and lower blocks enclose developed, hybrid and emergent countries, respectively.

the temporal causality. So, in order to gain a deeper insight, we analyze the localization of each country for this optimal situation in Fig. 2. The countries are labeled by the corresponding numbers listed in Table 1. We find a group of ten countries (Argentina, Austria, Brazil, Canada, China, Greece, Hong Kong, Singapore, Taiwan and Turkey) in an intermediate situation. So, we propose a more refined ranking of the stock markets. We divide all the countries under analysis in three different groups: developed, emergent and the intermediate group designed as *hybrid*, following the terminology used by Matos et al. [25]. Please see these three different blocks in Fig. 2.<sup>2</sup> The proposed hybrid group, with intermediate entropy and complexity values, is composed by five developed and five emergent stock markets according to the MSCI criterion. So, the first five were relegated to the intermediate group and the second five were promoted with our physical approach. The emergent countries that are included in the intermediate case are evolving very fast and their markets have become more efficient over time. On the other hand, the developed economies that comprise the intermediate case are relatively small if they are compared to other developed economies such as the US, Germany, Japan and United Kingdom. Therefore, there is some room to define an intermediate group that may be evolving or involving over time. See Ref. [13] for a discussion on evolution of financial markets over time.

In order to show the robustness of these results, we have estimated the  $\mathcal{H}_S$  and  $\mathcal{C}_J$  for the associated shuffled and surrogated prices with the same embedding dimension ( $D = 6$ ) and time delay ( $\tau = 1$ ). In the shuffling procedure the data are put into random order, and thus all non-trivial temporal correlations are destroyed. Surrogated time series are obtained from the original by randomizing their phases in the Fourier space and the generated series result Gaussian. Ten different realizations of the shuffled and surrogated time series related to the daily prices were generated to reduce the statistical errors. The associated mean values of both quantifiers for these new generated time series are illustrated in Fig. 3. The  $\mathcal{H}_S$  and  $\mathcal{C}_J$  in the shuffled case collapse to values near to 0.98 and 0.05, respectively. For a perfect Gaussian random process (white noise) it is expected that  $\mathcal{H}_S = 1$  and  $\mathcal{C}_J = 0$ . We have also estimated these two quantifiers for ten different realizations of 3138 normally distributed pseudorandom numbers. Time series were generated by using the Matlab function *randn*. The mean values estimated for the  $\mathcal{H}_S$  and  $\mathcal{C}_J$  were  $0.981 \pm 0.002$  and  $0.047 \pm 0.005$ , respectively. Moreover, we have found that the values are closer to the ideal values (i.e.  $\mathcal{H}_S \rightarrow 1$  and  $\mathcal{C}_J \rightarrow 0$ ) when the length of the record is increased. Thus, the differences are due to finite-size effects. We verify in this way that the localizations obtained from real data are not obtained by chance. Moreover, the underlying correlations are significant and play a relevant role. In the surrogated case we have found that the  $\mathcal{H}_S$  increases and the  $\mathcal{C}_J$  diminishes for all countries. So, the surrogated time series move to regions of higher efficiency. However, the displacements obtained are notably smaller than in the shuffled case. From this analysis of the localizations of the associated shuffled and surrogated time series in the complexity-entropy causality plane, it is possible to conclude that the stock market inefficiency can be primarily attributed to the long-range correlations and secondarily to the broad fat-tail distributions.

It is important to note that we have only considered six consecutive days to build the ordinal patterns. However, the information related with the long-range correlations is provided by the permutation probability distribution associated to each financial index. Differences in the long-range correlations of time series translate to differences in their associated probability distribution. Moreover, it was shown theoretically and through numerical simulations that the normalized

<sup>2</sup> It should be stressed that the localization and size of the blocks were chosen by the authors. So, they are in some way arbitrary. A challenge for the future will be to define the limits between the different groups in a more systematic and objective approach.



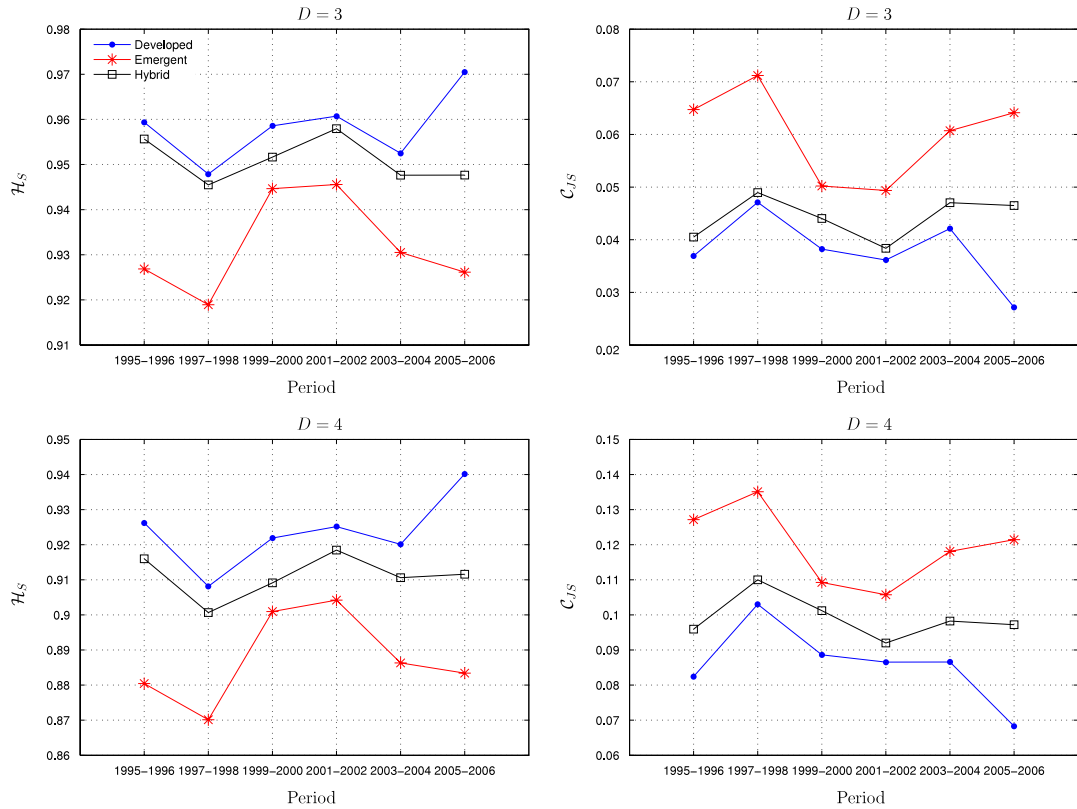
**Fig. 3.** Localization of the shuffled (upper plot) and surrogated (lower plot) stock markets in the complexity-entropy causality plane with embedding dimension  $D = 6$  and time delay  $\tau = 1$ . The shuffled and surrogated values are estimated averaging over ten different realizations. The maximum and minimum possible values of the intensive statistical complexity measure are also shown (segmented lines).

permutation entropy and the intensive SCM are able to distinguish stochastic processes with different long-range correlations, like the fractional Gaussian noise and fractional Brownian motion, for embedding dimensions  $D = 3, 4, 5$  and  $6$  [29,58,62].

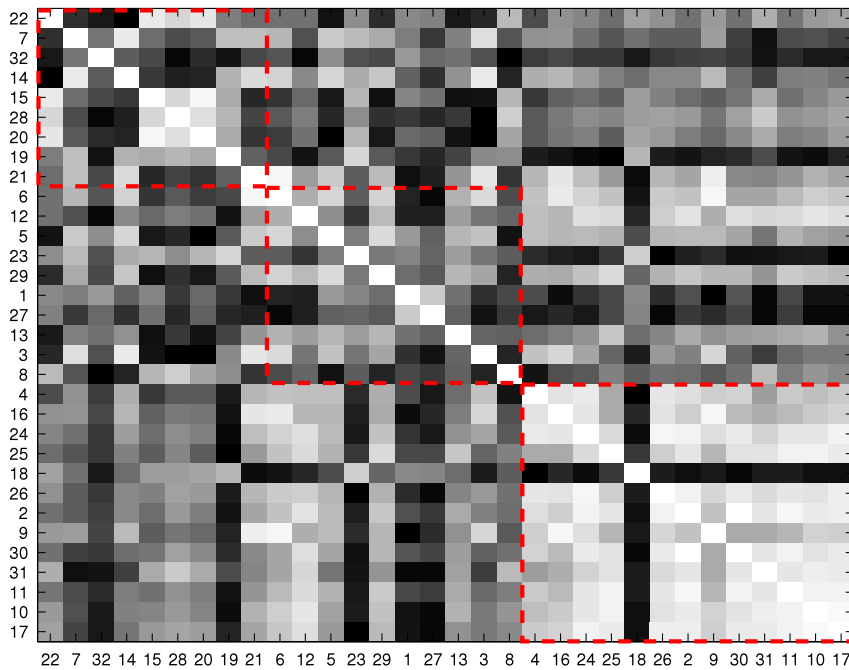
We have also analyzed the temporal evolution of the two quantifiers for the three groups: developed, emergent, and hybrid. The original time series were divided into six periods of 500 observations—each period corresponds with two years approximately. The average over all the developed, emergent and hybrid countries, respectively, for each one of these periods is plotted in Fig. 4. Observe that our classification is robust for different time periods. An important finding is that the hybrid group shares similar dynamics with developed economies; they seem to be highly correlated. An additional observation is that the emergent group appears to be more volatile, which is to be expected. A potential explanation for these results could be that the hybrid group is composed of stock markets that have become more integrated with world market, whereas emergent markets are more resilient and not highly integrated.

With the aim of testing whether the localization in the complexity-entropy causality plane is related with the linear correlations between the different countries, we have calculated the matrix of Pearson's correlation coefficients.<sup>3</sup> The absolute value of  $\rho_{ij}$ ,  $0 \leq |\rho_{ij}| \leq 1$  between countries  $i$  and  $j$ , was considered. Fig. 5 shows the gray-scale map of these linear cross-correlations. The countries are ordered in both axes following the complexity-entropy causality plane criterion. Obviously the map is symmetrical, centering around the diagonal. Observe that there are higher linear correlations (white

<sup>3</sup> Similar results were obtained for the Spearman's rank correlation coefficient.

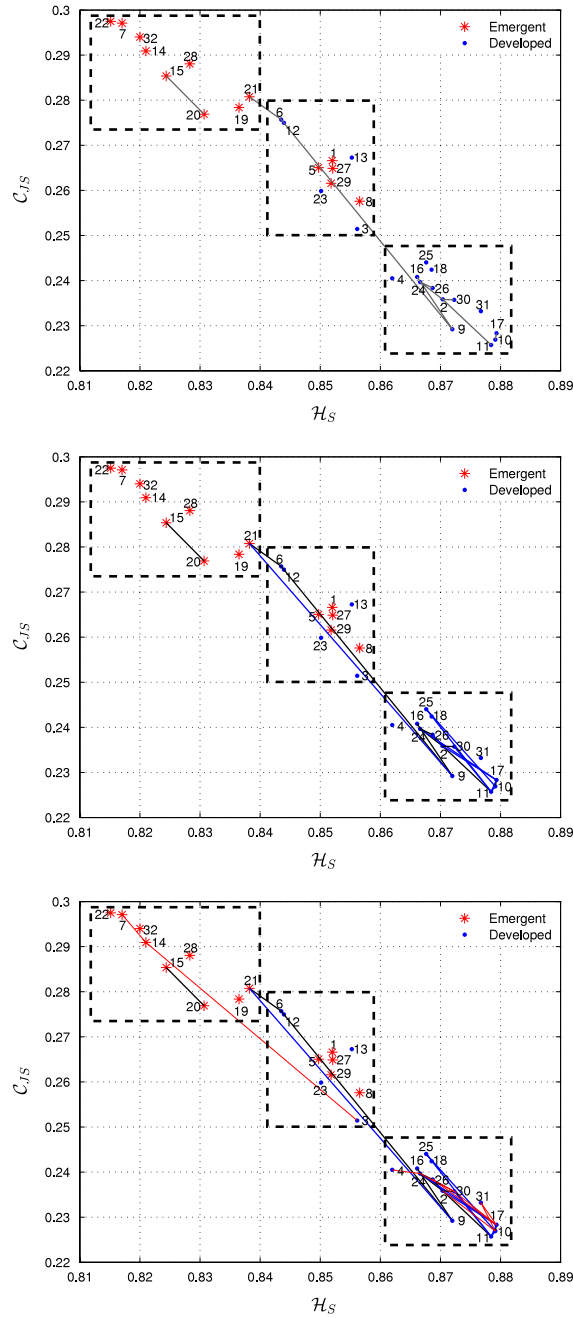


**Fig. 4.** Temporal evolution of the  $\mathcal{H}_S$  and  $\mathcal{C}_{JS}$  averaged over all the countries of each group for periods of two years approximately. Embedding dimensions  $D = 3$  (upper plots) and  $D = 4$  (lower plots), and time delay  $\tau = 1$  were considered.



**Fig. 5.** Gray-scale map of the cross-correlation coefficients (absolute value) between the different stock markets. White square means perfect linear correlation ( $\rho_{ij} = -1$  or  $\rho_{ij} = 1$ ) and black square means absence of linear correlation ( $\rho_{ij} = 0$ ). Numbers indicate stock markets listed in Table 1. They are ordered following the complexity-entropy causality plane criterion and the blocks dividing developed, hybrid and emergent countries are depicted.





**Fig. 6.** Evolution of the cross-correlations in the complexity-entropy causality plane. Top:  $0.95 \leq |\rho_{ij}| \leq 1$  (black links). Center:  $0.93 \leq |\rho_{ij}| < 0.95$  (blue links). Bottom:  $0.91 \leq |\rho_{ij}| < 0.93$  (red links).

squares) between the developed countries, in the lower right corner of the map. Only Japan, well-defined as developed following the complexity-entropy causality plane localization, has a low linear correlation with the countries of this cluster.<sup>4</sup> We guess that this particular behavior could be related to geographical influences. Note that Japan has higher correlations with Indonesia, Korea, Malaysia, Indonesia, Thailand and Philippines. It is also possible to identify another regional cluster, in the upper left corner of the map, composed of Indonesia, Malaysia and Thailand.

<sup>4</sup> Markets in different countries are open during different hours. In order to circumvent this lack of synchronicity we have also considered the two-day average prices. However, we have not found significant differences between the cross-correlation coefficients of the daily and two-day average prices. The results for the latter are available upon request.

We have also depicted the evolution of the linear correlations between the different countries within the complexity-entropy causality plane. Fig. 6 shows three different plots, corresponding to three different thresholds in the correlation strength. To be more specific, links of different colors (black, blue and red) are drawn for all the countries with linear correlations larger than 0.95, 0.93 and 0.91, respectively. We can clearly observe that some developed countries are strongly related to themselves, as if they were acting as a single market; moreover, if we lower the threshold, their influence spreads to less developed countries. However, there are other developed countries that are not correlated. Thus, we conclude the complexity-entropy causality plane provides additional information; i.e. linear but also nonlinear correlations for each stock market are quantified.

#### 4. Conclusions

We have shown that the complexity-entropy causality plane,  $C_{JS}$  versus  $\mathcal{H}_S$ , allows to distinguish the stage of stock market development. The differences between emergent and developed stock markets can be easily visualized with this statistical tool. The former have lower entropy and higher complexity values revealing the presence of significant time correlations and some degree of order. Moreover, a more refined classification of stock market dynamics is obtained. Analyzing the localizations of the associated shuffled and surrogated time series in this representation plane, it is possible to visualize and, consequently, quantify the influence of the two sources of inefficiency: long-range correlation and broad fat-tail distribution. We have found that, for the stock markets under analysis, the temporal correlations are the main factor of inefficiency. We stress the fact that the complexity-entropy causality plane is a model independent diagnostic tool, having more general applicability than other widely used alternatives like the Hurst parameter. So, we conclude that it can be considered a useful physical approach for an economy issue.

It will be interesting in the future to evaluate the influence of time delays  $\tau \neq 1$  in the results. Also the complexity-entropy causality plane localization of other financial time series, like foreign exchange rates and commodity markets, will be analyzed. Finally, further research will consider how our results are affected when other useful and well-known entropy definitions, like Renyi or Tsallis entropies, are applied.

#### Acknowledgements

Luciano Zunino and Osvaldo A. Rosso were supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. Benjamin M. Tabak gratefully acknowledges financial support from CNPq foundation. The opinions expressed in the paper do not necessarily reflect those of the Banco Central do Brasil. Darío G. Pérez was supported by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT, FONDECYT project No. 11060512), Chile, and partially by Pontificia Universidad Católica de Valparaíso (PUCV, Project No. 123. 788/2007), Chile. The authors are very grateful to the reviewers, whose comments and suggestions helped to improve an earlier version of this paper.

#### References

- [1] L. Bachelier, *Théorie de la spéculation*, Ph.D. Thesis, Sorbonne, Paris, 1900.
- [2] E.F. Fama, Efficient capital markets: A review of theory and empirical work, *J. Finance* 25 (1970) 383–417.
- [3] B. Mandelbrot, The variation of certain speculative prices, *J. Bus.* 36 (1963) 394–419.
- [4] M. Beben, A. Orłowski, Correlation in financial time series: Established versus emerging markets, *European Phys. J. B* 20 (2001) 527–530.
- [5] T. Di Matteo, T. Aste, M.M. Dacorogna, Scaling behaviors in differently developed markets, *Physica A* 324 (2003) 183–188.
- [6] T. Di Matteo, T. Aste, M.M. Dacorogna, Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development, *J. Bank. Finance* 29 (2005) 827–851.
- [7] C. Eom, S. Choi, G. Oh, W.-S. Jung, Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets, *Physica A* 387 (2008) 4630–4636.
- [8] C. Eom, G. Oh, W.-S. Jung, Relationship between efficiency and predictability in stock price change, *Physica A* 387 (2008) 5511–5517.
- [9] H.F. Coronel-Brizio, A.R. Hernández-Montoya, R. Huerta-Quintanilla, M. Rodríguez-Achach, Evidence of increment of efficiency of the Mexican Stock Market through the analysis of its variations, *Physica A* 380 (2007) 391–398.
- [10] R. Giglio, R. Matsushita, A. Figueiredo, I. Gleria, S.D. Silva, Algorithmic complexity theory and the relative efficiency of financial markets, *Europhys. Lett.* 84 (2008) 48005.
- [11] D.O. Cajueiro, B.M. Tabak, Evidence of long range dependence in Asian equity markets: The role of liquidity and market restriction, *Physica A* 342 (2004) 656–664.
- [12] D. Grech, Z. Mazur, Can one make any crash prediction in finance using the local Hurst exponent idea?, *Physica A* 336 (2004) 133–145.
- [13] D.O. Cajueiro, B.M. Tabak, The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient, *Physica A* 336 (2004) 521–537.
- [14] D.O. Cajueiro, B.M. Tabak, Ranking efficiency for emerging markets, *Chaos, Solitons & Fractals* 22 (2004) 349–352.
- [15] D.O. Cajueiro, B.M. Tabak, Ranking efficiency for emerging markets II, *Chaos, Solitons & Fractals* 23 (2005) 671–675.
- [16] L. Zunino, B.M. Tabak, D.G. Pérez, M. Garavaglia, O.A. Rosso, Inefficiency in Latin-American market indices, *European Phys. J. B* 60 (2007) 111–121.
- [17] D. Grech, G. Pamuła, The local Hurst exponent of the financial time series in the vicinity of crashes on the Polish stock exchange market, *Physica A* 387 (2008) 4299–4308.
- [18] Ł. Czarnecki, D. Grech, G. Pamuła, Comparison study of global and local approaches describing critical phenomena on the Polish stock exchange market, *Physica A* 387 (2008) 6801–6811.
- [19] K.E. Bassler, G.H. Gunaratne, J.L. McCauley, Markov processes, Hurst exponents, and nonlinear diffusion equations: With application to finance, *Physica A* 369 (2006) 343–353.
- [20] K. Matia, M. Pal, H. Salunkay, H.E. Stanley, Scale-dependent price fluctuations for the Indian stock market, *Europhys. Lett.* 66 (2004) 909–914.
- [21] G. Oh, S. Kim, C. Eom, Market efficiency in foreign exchange markets, *Physica A* 382 (2007) 209–212.
- [22] B.L. Lan, Y.O. Tan, Statistical properties of stock market indices of different economies, *Physica A* 375 (2007) 605–611.

- [23] A.A.G. Cortines, C. Anteneodo, R. Riera, Stock index dynamics worldwide: A comparative analysis, *European Phys. J. B* 65 (2008) 289–294.
- [24] L. Zunino, B.M. Tabak, A. Figliola, D.G. Pérez, M. Garavaglia, O.A. Rosso, A multifractal approach for the stock market inefficiency, *Physica A* 387 (2008) 6558–6566.
- [25] J.A. Matos, S.M. Gama, H.J. Ruskin, A.A. Sharkasi, M. Crane, Time and scale Hurst exponent analysis for financial markets, *Physica A* 387 (2008) 3910–3915.
- [26] W.A. Rizzo, The informational efficiency: The emerging markets versus the developed markets, *Appl. Econom. Lett.* 16 (2008) 485–487.
- [27] L. Zunino, M. Zanin, B.M. Tabak, D.G. Pérez, O.A. Rosso, Forbidden patterns, permutation entropy and stock market inefficiency, *Physica A* 388 (2009) 2854–2864.
- [28] M. Zanin, Forbidden patterns in financial time series, *Chaos* 18 (2008) 013119.
- [29] O.A. Rosso, H.A. Larrondo, M.T. Martín, A. Plastino, M.A. Fuentes, Distinguishing noise from chaos, *Phys. Rev. Lett.* 99 (2007) 154102.
- [30] G.A. Darbellay, D. Wuertz, The entropy as a tool for analysing statistical dependences in financial time series, *Physica A* 287 (2000) 429–439.
- [31] S.R. Bentes, R. Menezes, D.A. Mendes, Long memory and volatility clustering: Is the empirical evidence consistent across stock markets?, *Physica A* 387 (2008) 3826–3830.
- [32] Y.-C. Zhang, Toward a theory of marginally efficient markets, *Physica A* 269 (1999) 30–44.
- [33] A. Dionisio, R. Menezes, D.A. Mendes, An econophysics approach to analyse uncertainty in financial markets: An application to the Portuguese stock market, *European Phys. J. B* 50 (2006) 161–164.
- [34] D. Matesanz, G.J. Ortega, A (econophysics) note on volatility in exchange rate time series. Entropy as a ranking criterion, *Int. J. Mod. Phys. C* 19 (2008) 1095–1103.
- [35] S. Pincus, R.E. Kalman, Irregularity, volatility, risk, and financial market time series, *Proc. Natl. Acad. Sci.* 101 (2004) 13709–13714.
- [36] O. Kwon, J.-S. Yang, Information flow between stock indices, *Europhys. Lett.* 82 (2008) 68003.
- [37] W.A. Rizzo, The informational efficiency and the financial crashes, *Res. Int. Bus. Financ.* 22 (2008) 396–408.
- [38] P.W. Lamberti, M.T. Martín, A. Plastino, O.A. Rosso, Intensive entropic non-triviality measure, *Physica A* 334 (2004) 119–131.
- [39] R. López-Ruiz, H.L. Mancini, X. Calbet, A statistical measure of complexity, *Phys. Lett. A* 209 (1995) 321–326.
- [40] M.T. Martín, A. Plastino, O.A. Rosso, Statistical complexity and disequilibrium, *Phys. Lett. A* 311 (2003) 126–132.
- [41] M.T. Martín, A. Plastino, O.A. Rosso, Generalized statistical complexity measures: Geometrical and analytical properties, *Physica A* 369 (2006) 439–462.
- [42] D.P. Feldman, C.S. McTague, J.P. Crutchfield, The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing, *Chaos* 18 (2008) 043106.
- [43] X. Calbet, R. López-Ruiz, Tendency towards maximum complexity in a nonequilibrium isolated system, *Phys. Rev. E* 63 (2001) 066116.
- [44] O.A. Rosso, M.T. Martín, A. Figliola, K. Keller, A. Plastino, EEG analysis using wavelet-based information tools, *J. Neurosci. Meth.* 153 (2006) 163–182.
- [45] L. Zunino, D.G. Pérez, M.T. Martín, A. Plastino, M. Garavaglia, O.A. Rosso, Characterization of Gaussian self-similar stochastic processes using wavelet-based informational tools, *Phys. Rev. E* 75 (2007) 021115.
- [46] O.A. Rosso, H. Craig, P. Moscato, Shakespeare and other English Renaissance authors as characterized by Information Theory complexity quantifiers, *Physica A* 388 (2009) 916–926.
- [47] C. Bandt, B. Pompe, Permutation entropy: A natural complexity measure for time series, *Phys. Rev. Lett.* 88 (2002) 174102.
- [48] K. Keller, H. Laufer, Symbolic analysis of high-dimensional time series, *Int. J. Bifur. Chaos* 13 (2003) 2657–2668.
- [49] H.A. Larrondo, M.T. Martín, C.M. González, A. Plastino, O.A. Rosso, Random number generators and causality, *Phys. Lett. A* 352 (2006) 421–425.
- [50] A. Kowalski, M.T. Martín, A. Plastino, O.A. Rosso, Bandt-Pompe approach to the classical-quantum transition, *Physica D* 233 (2007) 21–31.
- [51] O. Rosso, R. Vicente, C. Mirasso, Encryption test of pseudo-aleatory messages embedded on chaotic laser signals: An information theory approach, *Phys. Lett. A* 372 (2007) 1018–1023.
- [52] K. Keller, M. Sinn, Ordinal analysis of time series, *Physica A* 356 (2005) 114–120.
- [53] M. Staniak, K. Lehnertz, Parameter selection for permutation entropy measurements, *Int. J. Bifur. Chaos* 17 (2007) 3729–3733.
- [54] M. Matilla-García, M. Ruiz Marín, Detection of non-linear structure in time series, *Econom. Lett.* 105 (2009) 1–6.
- [55] Y. Cao, W. Tung, J.B. Gao, V.A. Protopopescu, L.M. Hively, Detecting dynamical changes in time series using the permutation entropy, *Phys. Rev. E* 70 (2004) 046217.
- [56] Z. Liu, Measuring the degree of synchronization from time series data, *Europhys. Lett.* 68 (2004) 19–25.
- [57] X. Li, G. Ouyang, D.A. Richards, Predictability analysis of absence seizures with permutation entropy, *Epilepsy Res.* 77 (2007) 70–74.
- [58] L. Zunino, D.G. Pérez, M.T. Martín, M. Garavaglia, A. Plastino, O.A. Rosso, Permutation entropy of fractional Brownian motion and fractional Gaussian noise, *Phys. Lett. A* 372 (2008) 4768–4774.
- [59] O.A. Rosso, C. Masoller, Detecting and quantifying stochastic and coherence resonances via information-theory complexity measurements, *Phys. Rev. E* 79 (2009) 040106(R).
- [60] O.A. Rosso, C. Masoller, Detecting and quantifying temporal correlations in stochastic resonance via information theory measures, *European Phys. J. B* 69 (2009) 37–43.
- [61] C. Bandt, F. Shiha, Order patterns in time series, *J. Time Ser. Anal.* 28 (2007) 646–665.
- [62] O.A. Rosso, L. Zunino, D.G. Pérez, A. Figliola, H.A. Larrondo, M. Garavaglia, M.T. Martín, A. Plastino, Extracting features of Gaussian self-similar stochastic processes via the Bandt & Pompe approach, *Phys. Rev. E* 76 (2007) 061114.