Forbidden patterns, permutation entropy and stock market inefficiency

Luciano Zunino [a,b,c,*], Massimiliano Zanin [d], Benjamin M. Tabak [e,f], Darío G. Pérez [g], and Osvaldo A. Rosso [h,i]

[a] *Centro de Investigaciones Ópticas. C.C. 124 Correo Central. 1900 La Plata, Argentina.*

[b] *Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP). 1900 La Plata, Argentina.*

[c] *Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de La Plata. 1900 La Plata, Argentina.*

[d] *Universidad Autónoma de Madrid, 28049 Madrid, Spain.*

[e] *Banco Central do Brasil, SBS Quadra 3, Bloco B, 9 andar, DF 70074-900, Brazil.*

[f] *Universidade Catolica de Brasilia, Brasilia, DF, Brazil.*

[g] *Instituto de Física, Pontificia Universidad Católica de Valparaíso (PUCV). 23-40025 Valparaíso, Chile.*

[h] *Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, School of Electrical Engineering and Computer Science, The University of Newcastle. University Drive, Callaghan NSW 2308, Australia.*

[i] *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.*

Abstract

In this paper we introduce two new quantifiers for the stock market inefficiency: the number of forbidden patterns and the normalized permutation entropy. They are model-independent measures, thus they have more general applicability. We find robust evidence that degree of market inefficiency is positively correlated with the number of forbidden patterns and negatively correlated with the permutation entropy. Our empirical results suggest that these two physical tools are useful to discriminate the stage of stock market development and can be easily implemented.

PACS: 89.65.Gh (Economics; econophysics, financial markets, business and management), 05.45.Tp (Time series analysis), 89.75.Da (Systems obeying scaling laws), 89.20.-a (Interdisciplinary applications of physics)

KEYWORDS: Forbidden patterns, Permutation entropy, Stock market inefficiency

1 Introduction

The French mathematician Louis Bachelier, almost a century ago, modeled the market prices by what today is known as a random walk [1]. Such a walk is a random process where the increments are uncorrelated. Moreover, according to the weak-form version of the celebrated Efficient Market Hypothesis (EMH) [2], at any given time, the price of an asset fully reflects all available information. It should follow a martingale process in which each price change is unmodified by its predecessor. However, it is now widely known that empirical evidence suggests that even the most competitive markets are not strictly efficient. Price histories can be used to predict near future returns with a probability better than random chance.

A quantifier for the stock market inefficiency is of paramount importance for policy makers and regulators since it provides some guidance on appropriate policies that may be used to improve market efficiency [3,4], which by its turn reduces distortions in the economy [*]. Furthermore, it is crucial for the investment funds industry and portfolio and risk management purposes. Concepts like economic management, political stability and risk profile distinguish emerging and developed stock markets. However, these properties are very difficult to quantify in a precise way [5]. It was also shown that the liquidity and market capitalization play an important role in understanding the market inefficiency [6]. Very recently, a positive relationship between stock market inefficiency and predictability was experimentally assessed [7,8].

Several authors have proposed the Hurst parameter as a measure of the stock market inefficiency [7–16]. More precisely, they argue that this scaling exponent measures the long-range dependence exhibited by the stock market indices under analysis. Obviously, the existence of autocorrelation between distant observations violates the market efficiency because past prices can

* Corresponding author.
Email addresses: lucianoz@ciop.unlp.edu.ar (Luciano Zunino), massimiliano.zanin@hotmail.com (Massimiliano Zanin), benjamin.tabak@bcb.gov.br (Benjamin M. Tabak), dario.perez@ucv.cl (Darío G. Pérez), oarosso@fibertel.com.ar (Osvaldo A. Rosso).

[*] Market efficiency implies that firms are able to finance themselves using the correct cost of capital, that investors receive equilibrium returns, which are adjusted for the risk they incur, and that portfolio and risk managers can price the risks of their portfolios and make better decisions. Therefore, market efficiency should be related to lower distortions in the economy.

help to predict future prices allowing the possibility of arbitrage opportunities. However, Bassler *et al.* [17] 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. The Markov processes, by construction, have no memory and, in this case, the Hurst parameter is associated to the nonstationary increments of the process. Thus, they conclude that for an appropriately evidence for autocorrelations in data it is mandatory to show that the increments are stationary besides an empirical measurement or theoretical prediction of the Hurst parameter. The motivation of the present effort is just to introduce an alternative model-independent quantifier, of more general applicability, for the stock market inefficiency in order to be able to discriminate the stage of stock market development.

Important regularities between financial and physical data were found [18,19]. Consequently, concepts and method of statistical physics are increasingly being applied to economics. It should be stressed that different approaches have been recently introduced to rank stock markets in order to distinguish between emerging and developed economies [20–22]. Furthermore, 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 constraints on the theoretical probability distribution [23,24]. If prices were a pure random walk, the variations would be a completely uncorrelated string of numbers. We would say that such a string of variables 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 called negative entropy, i.e., the price variation's entropy with reference to its maximal value, can be taken as a measure of predictability and, consequently, of market inefficiency [25].

In order to evaluate the entropy of a time series an associated probability distribution must be determined. Many methodologies have been proposed for this purpose. We can mention, among others, procedures based on Fourier analysis [26], symbolic analysis [27], amplitude statistics [28] and wavelet transform [29]. Their applicability depends on particular characteristics of the data like stationarity, length of the time series, variation of the parameters, level of noise contamination, etc. In all these cases the global aspects of the dynamics can be in some way captured, but the different approaches are not equivalent in their ability to discern all the relevant physical details. The adequate way of picking up the probability distribution associated to a time series is a fundamental problem. Rosso *et al.* [30] have recently shown that improvements in the results can be expected if the underlying probability distribution is obtained taking into account the time causality of the system's dynamics. Specifically, it was found that different information measures allow to distinguish between chaotic and stochastic dynamics when causal information is incorporated into the scheme to extract the associated probability distribution. Bandt & Pompe [31] 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 attractor 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. A notable result from the Bandt & Pompe approach is a notorious improvement in the performance of the information quantifiers obtained using the probability distribution generated by their algorithm [32–37]. It should be also mentioned that there is an error associated to this way of estimating the probability distribution. The bias of the estimated entropy by employing the usual histogram method, based on counting the relative frequencies of the time series values within each subinterval, has been already accomplished [38]. However, the estimated entropy following the BPM, popularly known as permutation entropy, quantifies the diversity of possible ordering of the time series values [39] and its associated bias has not been calculated yet.

In this paper, we use the forbidden patterns and permutation entropy concepts to quantify the stock market inefficiency. The existence of missing sequences in a given time series is a recently proposed tool to discriminate random from deterministic time series [40]. On the other hand, the permutation entropy is just the celebrated Shannon entropic measure evaluated using the BPM to extract the associated probability distribution. Both are model-independent measures, avoiding restrictive assumptions on the probability distribution generating the data. Therefore, they can distinguish classes of systems for a wide variety of data, applications, and models[†] . We should remark that other entropies, like the Approximate Entropy (ApEn) [44,45], the Renyi and Tsallis ones [24] and the Shannon local entropy [46] have been proposed to quantify different aspect of financial time series. Careful reading of these papers is strongly suggested. In order to empirically test the hypothesis that the degree of market inefficiency is positively correlated with the number of forbidden patterns and negatively correlated with the permutation entropy we analyze the stock market indices of 32 countries. We find that these two physical tools are able to characterize and discriminate financial time series.

The dichotomy developed/emerging stock market is widely used within the finance field in order to segment markets that have different characteristics. In general, emerging markets have a fast growth pace, but still lack well-developed institutions and are seen as an investment class with higher risk,

[†] The permutation entropy has recently been applied to characterize a widely used price model like the fractional Brownian motion (fBm) [41–43].

and higher returns. For example, recent crisis have hit hard emerging markets and in some cases have ceased capital inflows for a while. Understanding the different characteristics of developed and emerging stock markets is crucial for the design of investment policies. We expect emerging stock markets to be more inefficient than developed markets due to the institutional and financial constraints that emerging markets face. Nonetheless, this is an empirical question that we seek to address in this paper.

The reminder of the paper is organized as follows. In the following section we describe the Bandt and Pompe method, and we introduce the concepts of forbidden patterns and permutation entropy. The data sets analyzed are detailed in Sec. 3. In Sec. 4 we present the empirical results obtained for the different stock market indices under study and several statistical tests are introduced to check whether the number of forbidden patterns and the permutation entropy are associated to the degree of market inefficiency. Finally, we summarize the finding of this paper in Sec. 5.

2 Bandt & Pompe method

Given a time series $\{x_t : t = 1, \ldots, M\}$, an embedding dimension $D > 1$, and a time delay τ, consider the ordinal patterns of order D [31,32,47] generated by

$$s \mapsto \left(x_{s-(D-1)\tau}, x_{s-(D-2)\tau}, \ldots, x_{s-\tau}, x_s\right). \tag{1}$$

To each time s we are assigning a D-dimensional vector that results from the evaluation of the time series at times $s, s-\tau, \ldots, s-(D-1)\tau$. 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 $\pi = (r_0, r_1, \cdots, r_{D-1})$ of $(0, 1, \cdots, D-1)$ defined by

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

In this way the vector defined by Eq. (1) is converted into a unique symbol π. Let us make a numerical example; we start with the time series $\{1, 3, 5, 4, 2, 5, \ldots\}$, and we set the embedding dimension $D = 4$. 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. (2) 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. 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 π_i of order D, their associated relative frequencies can be naturally computed by

$$p(\pi_i) = \frac{\sharp\{s|s \geq 1 + (D-1)\tau, s \text{ has ordinal pattern } \pi_i\}}{M - (D-1)\tau}, \tag{3}$$

where $\sharp$ is the cardinality of the set—roughly speaking, the number of elements in it. Thus, an ordinal patterns probability distribution $P = \{p(\pi_i), i = 1, \ldots, D!\}$ is obtained from the time series. To determine $p(\pi_i)$ exactly an infinite time series should be considered, taking $M \to \infty$ 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 [31].

The advantages of the BPM reside in its simplicity, its robustness, and its invariance with respect to nonlinear monotonous transformations. Also, this method provides an extremely fast computational algorithm. It can be applied to any type of time series (regular, chaotic, noisy, or experimental) [31], and does not require large amount of samples like others indicators (e.g. fractal analysis). Of course, the embedding dimension D plays an important role for the evaluation of the appropriate probability distribution, since D determines the number of accessible states, $D!$, and tells us about the necessary length M of the time series needed in order to work with a reliable statistics. In particular, Bandt & Pompe suggest for practical purposes to work with $3 \leq D \leq 7$. Concerning this last point in all calculations reported here the condition $M \gg D!$ is satisfied [36]. Another similar conditions can be found like $M \geq 5D!$ [48] or $M > (D+1)!$ [49]. The probability distribution P is obtained once we fix the embedding dimension D and the time delay τ. It should be stressed that these parameters have to be selected. However, Matilla-García has provided a mechanism by which the selection of D can be done automatically [48]: for a given data set of M observations, the embedding dimension should be the maximum D that satisfies $M \geq 5D!$. Hereafter, we always consider $\tau = 1$. Thus, in this work the influence of different time delays is not analyzed.

2.1 Forbidden patterns

The study of the order patterns has been proposed as a technique for evaluating the determinism of a given time series [40]. Every group of D adjacent and overlapped values of the time series has an associated permutation π_i according to Eq. (2). This permutation will be one of the $D!$ possible permutations. If the series has a random behavior, any permutation can appear. Moreover, the probability distribution of the ordinal patterns should be uniform because any permutation has the same probability of appearance when the dataset is long enough to exclude statistical fluctuations. Nevertheless, when the series corresponds to a chaotic variable, there are some patterns that cannot be found due to the underlying deterministic structure. They are the so-called forbidden patterns. The existence of forbidden patterns indicates an underlying deterministic behavior. It is worth mentioning that the occurrence of this forbidden patterns is intrinsic to its chaotic nature and adding samples to the series does not reduce their number. In Fig. 1 the number of forbidden patterns is plotted as a function of the number of samples for a random time series and for an equivalent time series generated by a logistic map, which is obtained from $x_{n+1} = 4x_n(1 - x_n)$ and $0 \leq x_0 \leq 1$. At first glance, you can conclude that the deterministic time series generates a higher number of forbidden patterns and that it is independent of the series length, revealing its chaotic nature. Moreover, it was shown that forbidden patterns can also distinguish chaos from randomness in finite time series contaminated with observational white noise [49]. Zanin has recently studied the appearance of forbidden patterns in different financial time series finding evidence of deterministic forces in the medium and long term [50]. Additionally, he has found that the evolution of the number of forbidden patterns (NFP) could be an appropriate tool to quantify the randomness of certain time periods within the financial series.

2.2 Permutation entropy

The foundation of the normalized permutation entropy (NPE) is to study the overall statistics of the ordinal patterns defined by Eq. (2). Furthermore, it is just the normalized Shannon entropy associated to the probability distribution $P = \{p(\pi_i), i = 1, \ldots, D!\}$

$$\mathcal{H}_S[P] = S[P]/S_{\max} = \left[-\sum_{i=1}^{D!} p(\pi_i)\ \ln(p(\pi_i)) \right] /S_{\max}, \tag{4}$$

where $S_{\max} = \ln D!$, $(0 \leq \mathcal{H}_S \leq 1)$—$S$ stands for Shannon entropy. The highest value, $\mathcal{H}_S[P] = 1$, is attained for a totally random sequence where all symbols appear with the same probability. On the other hand, the lowest one, $\mathcal{H}_S[P] = 0$, corresponds with an increasing or decreasing sequence. It has been recently proposed a non-parametric test for independence based on symbolic dynamics [48]. In this work, Matilla-García has shown that this test is closely related to the permutation entropy concept. Dependence of the data generating process introduces patterns in the time series. Hence, the permutation entropy decreases because the ordinal patterns are distant from sharing the same probability. It is very important to observe that the NFP analysis is concerned with the ordinal patterns themselves whereas the NPE is related with their associated probability distribution P.

3 Data

In this paper we analyze the NFP and NPE of 32 equity indices and returns for different countries. Let $x(t)$ be the equity index of a stock on a time t, the equity index returns, r_t, are calculated as its logarithmic difference, $r_t = \log(x(t+1)/x(t))$. All data were collected from the Bloomberg database. The codes and names of these indices are presented in Table 1. We employ daily data beginning in 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. 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 emerging stock markets are considered. This classification is obtained following the Morgan Stanley Capital Index (MSCI) methodology to define developed and emerging stock markets (http://www.mscibarra.com). This is particularly useful as it allows a direct comparison of classes of countries.

4 Empirical results

In Tables 2 and 3 we report the NFP and NPE values, respectively, of the developed and emerging daily equity indices and returns detailed previously. In our analysis we have selected embedding dimensions $D = 6$ [‡] for the NFP and $D = 4$, $D = 5$ and $D = 6$ for the NPE. Observe that we have on average 3138 observations. Thus, according to the condition introduced at the end of Sec. 2, the highest possible embedding dimension is 6 ($3138 \gg 6! = 720$). In the case of returns there are very few forbidden patterns and the NPE takes values very near to 1, which is the value expected for a completely randomized time series. It is widely known that returns time series are much less correlated than prices time series. Zunino *et al.* [43] have worked out theoretical curves for the normalized permutation entropy of the fractional Brownian motion and its associated noise, the fractional Gaussian noise. In this work the authors found that the processes have much smaller NPE than the noises. Moreover, it was shown that the NPE for the noises are very near to 1—see Figs. 3 and 4 of Ref. [43].

In order to show a robustness of results, we have estimated the NFP and NPE

‡ For embedding dimensions $D = 4$ and $D = 5$ we have $4! = 24$ and $5! = 120$ possible permutations, respectively. Comparing these numbers with our time series lengths we conclude that there is a very low probability to find forbidden patterns. Then, these embedding dimensions are not considered in the forbidden patterns analysis.

for the shuffled prices and returns with embedding dimension $D = 6$. In the shuffling procedure the data are put into random order. So, all non-trivial temporal correlations are destroyed. The results are listed in Table 4. Ten different realizations of the shuffled time series associated to the daily prices and returns were generated in order to reduce the statistical errors. We have found that for both, prices and returns time series, the NFP and the NPE show significant differences compared with those associated to the shuffled counterparts [§]. However, as we have obtained better results for prices than returs, hereafter only the former will be considered in the analysis.

With the aim of testing whether the NFP is associated to the degree of market inefficiency we run an Ordinary Least Squares (OLS) regression with the NFP for $D = 6$, detailed in the second column of Table 2, as independent variable and a dummy assuming value one if the index being evaluated is that from a developed market and zero otherwise (emerging market) [¶]. The regression has the following form:

$$\gamma = \alpha_0 + \beta_0 \mathrm{NFP} + \epsilon, \tag{5}$$

where γ is a dummy for market development, α_0 and β_0 are parameters to be estimated and ϵ is an error term, assumed to be normally distributed. The β_0 coefficient is expected to be negative and statistically significant if emerging markets have a higher NFP, on average. The estimated coefficients are $\alpha_0 = 2.749 \pm 0.340$ and $\beta_0 = -0.014 \pm 0.002$. Both coefficients are statistically significant at the 1% level. The coefficient β_0 has the correct sign and one can infer that emerging markets have a higher NFP. The adjusted R^2 is equal to 57.36%, which suggests that the dummy variable is able to explain a large portion of the variability in the NFP. In order to check these results we also compare the median of forbidden patterns for developed and emerging markets indices using the Mann-Whitney non-parametric test. The medians for developed and emerging markets are 133 and 182, respectively. The Mann-Whitney statistic is 3.93 and we can reject the null hypothesis that these medians are equal at the 1% significance level.

We also employ a binary dependent variable model—see [51] for further details. The dependent variable in our model is a dummy variable that takes value one if the stock market under analysis is a developed market and zero if it is an emerging market developed market, as in the previous case. We employ three specifications for this dependent variable model assuming a logistic, normal and extreme value distributions for the errors in the model. Therefore, we test

[§] By employing the Mann-Whitney U test we have estimated the $p-$value associated to the null hypothesis that the values come from distributions with equal medians. The $p-$values obtained for the four comparisons: i) NFP for prices, ii) NFP for returns, iii) NPE for prices and iv) NPE for returns were 6.4186 10^{-12}, 1.4637 10^{-5}, 6.5113 10^{-12} and 1.6819 10^{-5}, respectively.

[¶] Results are robust to heterocedascity.

the results using the following functional forms, respectively:

$$Pr(x = 1|\mathbf{y}, \boldsymbol{\beta_1}) = G(\mathbf{y}'\boldsymbol{\beta_1}), \tag{6}$$

with G the cumulative distribution function of the standard normal distribution,

$$Pr(x = 1|\mathbf{y}, \boldsymbol{\beta_2}) = \frac{e^{\mathbf{y}'\boldsymbol{\beta_2}}}{1 + e^{\mathbf{y}'\boldsymbol{\beta_2}}}, \tag{7}$$

and

$$Pr(x = 1|\mathbf{y}, \boldsymbol{\beta_3}) = exp(-e^{(-\mathbf{y}'\boldsymbol{\beta_3})}). \tag{8}$$

The explanatory variable $\mathbf{y}$ is the NFP. If the coefficients $\boldsymbol{\beta}$ are negative and statistically significant then one can conclude that increasing the NFP will increase the probability of observing an emerging market. The estimated coefficients are $\beta_1 = -0.058 \pm 0.018$, $\beta_2 = -0.100 \pm 0.034$, and $\beta_3 = -0.063 \pm 0.019$ for the normal, logistic and extreme value specifications, respectively[‖]. In all cases the NFP is statistically significant at the 1% level. The McFadden R^2 are 54.81%, 54.10% and 50.95% for these three models, which implies a reasonable fit.

We have also evaluated the significance of the NPE for prices with $D = 4$, $D = 5$ and $D = 6$, detailed in the second, third and fourth column of Table 3, respectively, in assessing the stage of stock market development. Results are summarized in Table 5 and they suggest that developed markets have larger values of entropy, and therefore, are less predictable than emerging markets. Moreover, as we can easily conclude, the results are not dependent on the choice of the embedding dimension and are very robust. It should be stressed, however, that the best results are obtained for $D = 6$.

The Pearson's linear correlation coefficient between the NFP and NPE for $D = 6$ is -93.29%, whereas the Spearman's rank correlation coefficient is -92.34%. These correlations are both significant at the 1% level and imply that the information contained in these indices is similar. We orthogonalize the information in the NPE relative to the NFP and find that the remainder has no additional explanatory power for the stage of market development.

Therefore, we conclude that either the NFP or NPE may be used to assess the stage of stock market development. These results are robust to three types of modeling: an OLS regression, binary variable regressions and non-parametric median test.

We have also found a negative relationship between the degree of country development, measured by its economic size through the Gross National Income

‖ We estimate all models including a constant.

(GNI) per capita[**], and the NFP with embedding dimension $D = 6$. Figure 2 shows this relationship for prices and returns. The Spearman's rank correlation coefficients ρ[††] are -71.62% and -47.92%, respectively. We can observe that more developed countries tend to have lower NFP. As we have previously mentioned the NFP has worse results for returns because in this case there are very few forbidden patterns. Furthermore, a positive relationship is found between the GNI and the NPE for price and return time series with $D = 4$ ($\rho = 70.64\%$ for prices and $\rho = 62.32\%$ for returns), $D = 5$ ($\rho = 73.64\%$ for prices and $\rho = 61.03\%$ for returns) and $D = 6$ ($\rho = 74.23\%$ for prices and $\rho = 67.23\%$ for returns) as you can conclude from Fig. 3. The NPE has also better results for prices than returns and, again, the best results are obtained for $D = 6$. These results reinforce our previous findings and suggest that these quantifiers may be useful in assessing the stage of development of these markets.

5 Conclusions

In this paper we have introduced two new potentially useful quantifiers for the stock market inefficiency: the number of forbidden patterns and the normalized permutation entropy. They are model-independent in contrast with the inefficiency measures based on the Hurst parameter. Moreover, they can be easily implemented. Our findings show that these two physical concepts are helpful to discriminate stock market dynamics. Developed markets have lower NFP and higher NPE allowing to claim that they are less deterministic and less correlated. The slower reaction to new information, mostly associated to emerging stock markets, introduces characteristic patterns in the time series. Our quantifiers are able to detect the formation of these patterns. It should be stressed that the results obtained for price time series are better than those for return time series.

Further analysis for different values of the time delay τ and with another entropy measures, like the Tsallis permutation entropy recently introduced in [52], are planned. It was also recently shown that the permutation entropy is able to predict epileptic seizures [53]. Having into account the parallelism

[**] The GNI per capita measures average wealth of a particular country and therefore it can be considered as a proxy of the stage of development. The GNI data were obtained from the World Bank website and corresponds to the year 2006. As the GNI does not change much over time we consider that this value can be associated to the 1995-2007 period.

[††] The Spearman's rank correlation coefficient is of more general applicability because it is not restricted to normal distributions like the Pearson's linear correlation coefficient.

between living organisms and socio-economic systems [54], we infer that the permutation entropy could be a suitable statistical tool to predict financial crashes. This conclusion requires somewhat more elaboration and will be the subject of future work.

Acknowledgements

Luciano Zunino was 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. Osvaldo A. Rosso gratefully acknowledges support from Australian Research Council (ARC) Centre of Excellence in Bioinformatics, Australia. The authors are very greateful 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] N. Groenewold, S. H. K. Tang, Y. Wu, The efficiency of the Chinese stock market and the role of the banks, Journal of Asian Economies 14 (2003) 593–609.

[4] N. T. Laopodis, Fiscal policy and stock market efficiency: Evidence for the United States, The Quarterly Review of Economics and Finance, in press.

[5] 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.

[6] 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.

[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] D. Grech, Z. Mazur, Can one make any crash prediction in finance using the local Hurst exponent idea?, Physica A 336 (2004) 133–145.

[10] 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.

[11] D. O. Cajueiro, B. M. Tabak, Ranking efficiency for emerging markets, Chaos, Solitons & Fractals 22 (2004) 349–352.

[12] D. O. Cajueiro, B. M. Tabak, Ranking efficiency for emerging markets II, Chaos, Solitons & Fractals 23 (2005) 671–675.

[13] 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. Banking & Finance 29 (2005) 827–851.

[14] L. Zunino, B. M. Tabak, D. G. Pérez, M. Garavaglia, O. A. Rosso, Inefficiency in Latin-American market indices, Eur. Phys. J. B 60 (2007) 111–121.

[15] 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.

[16] 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.

[17] 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.

[18] H. E. Stanley, L. A. N. Amaral, D. Canning, P. Gopikrishnan, Y. Lee, Y. Liu, Econophysics: Can physicists contribute to the science of economics?, Physica A 269 (1999) 156–169.

[19] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley, Econophysics: financial time series from a statistical physics point of view, Physica A 279 (2000) 443–456.

[20] B. L. Lan, Y. O. Tan, Statistical properties of stock market indices of different economies, Physica A 375 (2007) 605–611.

[21] A. A. G. Cortines, C. Anteneodo, R. Riera, Stock index dynamics worldwide: a comparative analysis, Eur. Phys. J. B 65 (2008) 289–294.

[22] 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.

[23] G. A. Darbellay, D. Wuertz, The entropy as a tool for analysing statistical dependences in financial time series, Physica A 287 (2000) 429–439.

[24] 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.

[25] Y.-C. Zhang, Toward a theory of marginally efficient markets, Physica A 269 (1999) 30–44.

[26] G. E. Powell, I. C. Percival, A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems, J. Phys. A: Math. Gen. 12 (1979) 2053–2071.

[27] C. S. Daw, C. E. A. Finney, E. R. Tracy, A review of symbolic analysis of experimental data, Rev. Sci. Instrum. 74 (2003) 915–930.

[28] L. D. Micco, C. M. Gonzalez, H. A. Larrondo, M. T. Martin, A. Plastino, O. A. Rosso, Randomizing nonlinear maps via symbolic dynamics, Physica A 387 (2008) 3373–3383.

[29] O. A. Rosso, M. L. Mairal, Characterization of time dynamical evolution of electroencephalographic epileptic records, Physica A 312 (2002) 469–504.

[30] O. A. Rosso, H. A. Larrondo, M. T. Martín, A. Plastino, M. A. Fuentes, Distinguishing noise from chaos, Phys. Rev. Lett. 99 (2007) 154102.

[31] C. Bandt, B. Pompe, Permutation entropy: A natural complexity measure for time series, Phys. Rev. Lett. 88 (2002) 174102.

[32] K. Keller, H. Lauffer, Symbolic analysis of high-dimensional time series, Int. J. Bifurcation & Chaos 13 (2003) 2657–2668.

[33] 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.

[34] H. A. Larrondo, C. M. González, M. T. Martín, A. Plastino, O. A. Rosso, Intensive statistical complexity measure of pseudorandom number generators, Physica A 356 (2005) 133–138.

[35] 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.

[36] A. Kowalski, M. T. Martín, A. Plastino, O. A. Rosso, Bandt-Pompe approach to the classical-quantum transition, Physica D 233 (2007) 21–31.

[37] 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.

[38] F. Bazsó, L. Zalányi, A. Petróczi, Accuracy of joint entropy and mutual information estimates, Neural Networks, Proceedings of IEEE International Joint Conference on Neural Networks 4 (2004) 2843–2846.

[39] J. M. Amigó, M. B. Kennel, Topological permutation entropy, Physica D 231 (2007) 137–142.

[40] J. M. Amigó, L. Kocarev, J. Szczepanski, Order patterns and chaos, Phys. Lett. A 355 (2006) 27–31.

[41] I. Simonsen, K. Sneppen, Profit profiles in correlated markets, Physica A 316 (2002) 561–567.

[42] 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.

[43] 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.

[44] S. Pincus, R. E. Kalman, Irregularity, volatility, risk, and financial market time series, Proc. Natl. Acad. Sci. 101 (2004) 13709–13714.

[45] G. Oh, S. Kim, C. Eom, Market efficiency in foreign exchange markets, Physica A 382 (2007) 209–212.

[46] W. A. Risso, The informational efficiency and the financial crashes, Research in International Business and Finance 22 (2008) 396–408.

[47] K. Keller, M. Sinn, Ordinal analysis of time series, Physica A 356 (2005) 114–120.

[48] M. Matilla-García, A non-parametric test for independence based on symbolic dynamics, Jounal of Economic Dynamics & Control 31 (2007) 3889–3903.

[49] J. M. Amigó, S. Zambrano, M. A. F. Sanjuán, True and false forbidden patterns in deterministic and random dynamics, Europhys. Lett. 79 (2007) 50001.

[50] M. Zanin, Forbidden patterns in financial time series, Chaos 18 (2008) 013119.

[51] W. Greene, Econometric Analysis, 6th Edition, New York: Prentice Hall, 2007.

[52] L. Zunino, D. G. Pérez, A. Kowalski, M. T. Martín, M. Garavaglia, A. Plastino, O. A. Rosso, Fractional Brownian motion, fractional gaussian noise, and Tsallis permutation entropy, Physica A 387 (2008) 6057–6068.

[53] X. Li, G. Ouyang, D. A. Richards, Predictability analysis of absence seizures with permutation entropy, Epilepsy Research 77 (2007) 70–74.

[54] W. Klonowski, E. Olejarczyk, R. Stepien, Epileptic seizures in economic organism, Physica A 342 (2004) 701–707.

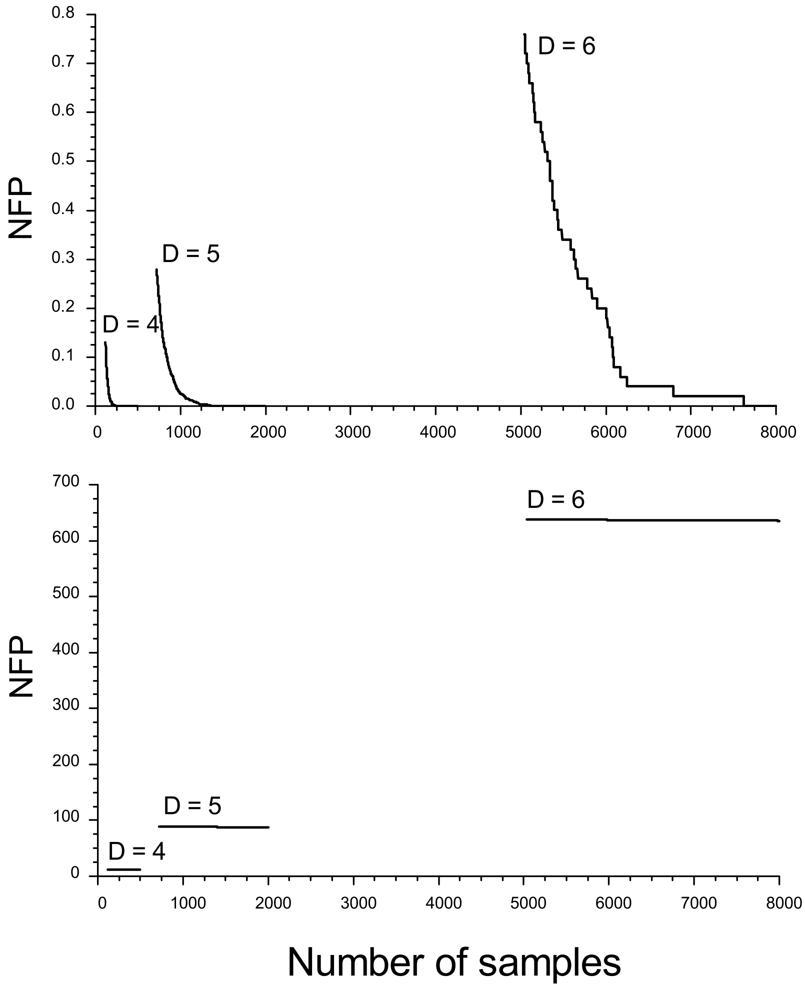

Fig. 1. Number of forbidden patterns for a random time series (top) and a deterministic time series generated by the logistic map (bottom) for embedding dimensions $D = 4$, $D = 5$ and $D = 6$.

Table 1
World stock indices.

Country	Bloomberg code	Index
Argentina	MERVAL	Argentina Merval Index
Australia	ASX	Australian Securities Exchange
Austria	ATX	Austrian Traded Index
Belgium	BEL20	BEL20 Index
Brazil	IBOV	Brazil Bovespa Stock Index
Canada	SPTSX	S&P/TSX Composite Index
Chile	IPSA	Chile Stock Market Select Index
China	SHCOMP	Shangai Stock Exchange Composite Index
Denmark	KFX	OMX Copenhagen 20 Index
France	CAC	CAC 40 Index
Germany	DAX	DAX Index
Greece	ASE	ASE General Index
Hong Kong	HSI	Hang Seng Index
India	SENSEX	Bombay Stock Exchange Sensitive Index
Indonesia	JCI	Jakarta Composite Index
Ireland	ISEQ	Irish Overall Index
Italy	MIB30	Milan MIB30 Index
Japan	TPX	Tokio Stock Price Index
Korea	KOSPI	Korea Composite Index
Malaysia	KLCI	Kuala Lumpur Composite Index
Mexico	MEXBOL	Mexico Bolsa Index
Philippines	PCOMP	Philippines Composite Index
Singapore	STI	Straits Times Index
Spain	IBEX	IBEX 35 Index
Sweden	OMX	OMX Stockholm 30 Index
Switzerland	SMI	Swiss Market Index
Taiwan	TWSE	Taiwan Taiex Index
Thailand	SET	Stock Exchange of Thailand Index
Turkey	XU100	Istambul Stock Exchange National 100 Index
United Kingdom	UKX	FTSE 100 Index
US	SPX	S&P 500 Index
Venezuela	IBVC	Venezuela Stock MKT Index

Table 2
Number of forbidden patterns for the world daily equity indices and returns with $\tau = 1$ and $D = 6$.

Developed markets	Prices	Returns
Australia	132	11
Austria	148	14
Belgium	123	11
Canada	189	12
Denmark	126	11
France	120	8
Germany	124	13
Greece	180	9
Hong Kong	179	18
Ireland	135	5
Italy	130	14
Japan	138	12
Singapore	158	12
Spain	130	6
Sweden	141	9
Switzerland	134	10
United Kingdom	130	13
US	126	5
Emerging markets	**Prices**	**Returns**
Argentina	178	10
Brazil	168	18
Chile	195	12
China	167	9
India	199	13
Indonesia	198	15
Korea	176	15
Malaysia	174	16
Mexico	190	18
Philippines	201	19
Taiwan	171	15
Thailand	186	12
Turkey	169	14
Venezuela	208	18

Table 3
Normalized permutation entropies for the world daily equity indices and returns with $\tau = 1$ and $D = 4$, $D = 5$ and $D = 6$.

Developed markets	Prices			Returns		
	$D = 4$	$D = 5$	$D = 6$	$D = 4$	$D = 5$	$D = 6$
Australia	0.929	0.903	0.870	0.999	0.996	0.981
Austria	0.924	0.894	0.856	0.999	0.996	0.981
Belgium	0.923	0.896	0.862	0.999	0.995	0.981
Canada	0.910	0.880	0.844	0.996	0.993	0.979
Denmark	0.935	0.909	0.872	0.998	0.996	0.982
France	0.935	0.911	0.879	0.999	0.997	0.983
Germany	0.938	0.912	0.878	0.998	0.993	0.978
Greece	0.911	0.880	0.844	0.997	0.994	0.979
Hong Kong	0.931	0.896	0.855	0.997	0.993	0.977
Ireland	0.928	0.901	0.866	0.999	0.997	0.982
Italy	0.942	0.914	0.879	0.999	0.995	0.979
Japan	0.930	0.902	0.869	0.998	0.995	0.980
Singapore	0.917	0.886	0.850	0.998	0.995	0.981
Spain	0.930	0.901	0.867	0.999	0.996	0.983
Sweden	0.930	0.902	0.868	0.999	0.996	0.982
Switzerland	0.930	0.903	0.869	0.999	0.996	0.982
United Kingdom	0.932	0.906	0.872	0.999	0.996	0.982
US	0.938	0.911	0.877	0.997	0.994	0.979
Emerging markets	Prices			Returns		
	$D = 4$	$D = 5$	$D = 6$	$D = 4$	$D = 5$	$D = 6$
Argentina	0.922	0.890	0.852	0.996	0.993	0.978
Brazil	0.917	0.886	0.850	0.997	0.994	0.977
Chile	0.887	0.855	0.817	0.994	0.990	0.977
China	0.929	0.897	0.857	0.997	0.993	0.978
India	0.897	0.860	0.821	0.996	0.992	0.977
Indonesia	0.894	0.862	0.824	0.996	0.993	0.977
Korea	0.904	0.872	0.836	0.996	0.993	0.978
Malaysia	0.902	0.869	0.831	0.998	0.993	0.976
Mexico	0.906	0.874	0.838	0.995	0.991	0.975
Philippines	0.889	0.853	0.815	0.997	0.993	0.977
Taiwan	0.922	0.888	0.852	0.998	0.994	0.980
Thailand	0.904	0.868	0.828	0.998	0.994	0.978
Turkey	0.917	0.887	0.852	0.999	0.997	0.981
Venezuela	0.889	0.857	0.820	0.995	0.990	0.974

Table 4
Number of forbidden patterns and normalized permutation entropies for the shuffled world daily equity indices and returns with $\tau = 1$ and $D = 6$. Both quantifiers are estimated averaging over the ten shuffled realizations.

Developed markets	NFP		NPE	
	Prices	Returns	Prices	Returns
Australia	10.1	7.7	0.982	0.982
Austria	7.8	8.6	0.982	0.981
Belgium	9.0	8.9	0.982	0.982
Canada	7.4	7.4	0.982	0.983
Denmark	9.7	10.3	0.981	0.981
France	9.5	10.1	0.982	0.982
Germany	9.0	8.7	0.982	0.982
Greece	9.2	6.9	0.982	0.982
Hong Kong	11.6	9.9	0.981	0.981
Ireland	8.8	9.2	0.982	0.981
Italy	7.7	10.5	0.982	0.981
Japan	9.0	10.9	0.981	0.981
Singapore	8.5	7.8	0.982	0.981
Spain	7.1	8.5	0.983	0.982
Sweden	9.3	9.3	0.982	0.982
Switzerland	8.9	9.4	0.982	0.982
United Kingdom	8.4	8.9	0.982	0.982
US	10.1	8.1	0.981	0.982
Emerging markets	NFP		NPE	
	Prices	Returns	Prices	Returns
Argentina	8.3	9.2	0.982	0.982
Brazil	9.7	8.7	0.982	0.982
Chile	8.8	9.7	0.982	0.982
China	10.4	9.9	0.981	0.981
India	9.4	8.3	0.982	0.982
Indonesia	10.4	9.2	0.981	0.981
Korea	6.8	7.2	0.982	0.982
Malaysia	9.4	10.3	0.981	0.981
Mexico	8.8	8.5	0.982	0.982
Philippines	7.9	9.0	0.982	0.982
Taiwan	6.5	5.8	0.983	0.983
Thailand	10.4	9.5	0.981	0.982
Turkey	10.5	9.0	0.981	0.982
Venezuela	10.4	12.1	0.982	0.981

Table 5
Results for the OLS regression, binary variable regressions (normal, logistic and extreme value specifications) and non-parametric median test between the normalized permutation entropies for different embedding dimensions (4, 5 and 6) and a dummy variable for emerging $\times$ developed markets. The symbol * stands for statistical significance at the 1% level.

	Embedding Dimension	Coefficient	Std.Error	Adj. R^2
OLS Regressions	4	22.76*	4.00	50.28%
	5	20.62*	3.26	55.73%
	6	19.42*	2.98	57.16%
Normal Specification	4	95.19*	29.24	48.07%
	5	96.41*	30.80	55.05%
	6	98.80*	34.15	57.69%
Logistic Specification	4	158.76*	52.83	47.18%
	5	162.18*	55.86	53.97%
	6	167.11*	61.35	56.59%
Extreme Value Specification	4	111.17*	36.57	47.46%
	5	110.58*	37.38	53.50%
	6	112.94*	40.85	55.69%
		Mann-Whitney	Developed	Emerging
Comparison of Median Entropy	4	4.08*	0.93	0.90
	5	4.12*	0.90	0.87
	6	4.16*	0.87	0.83

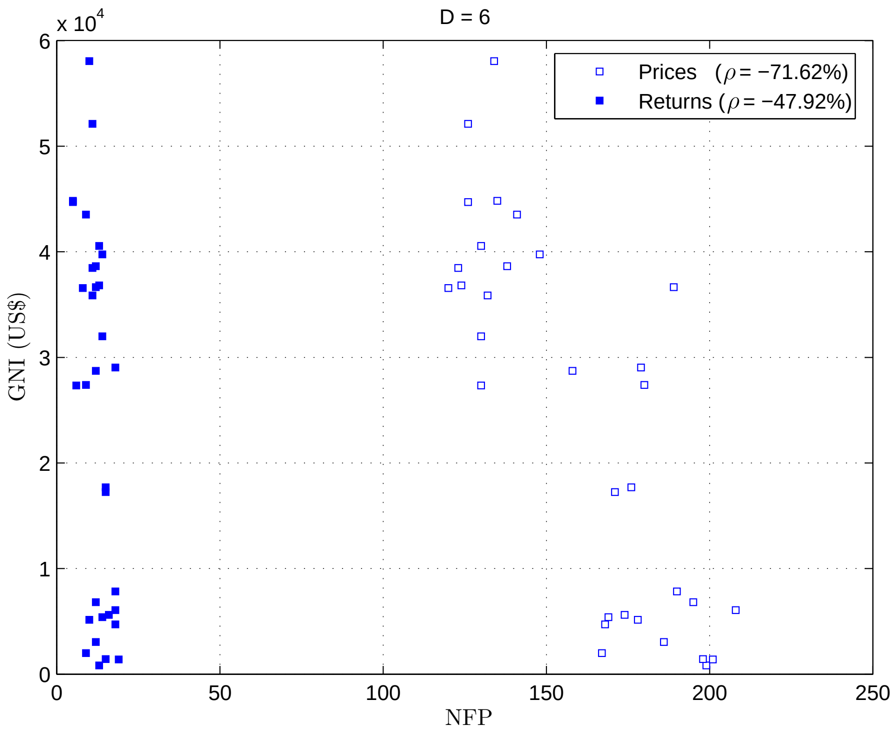

Fig. 2. The Gross National Income (GNI) per capita as a function of the NFP with embedding dimension $D = 6$ for prices and returns.

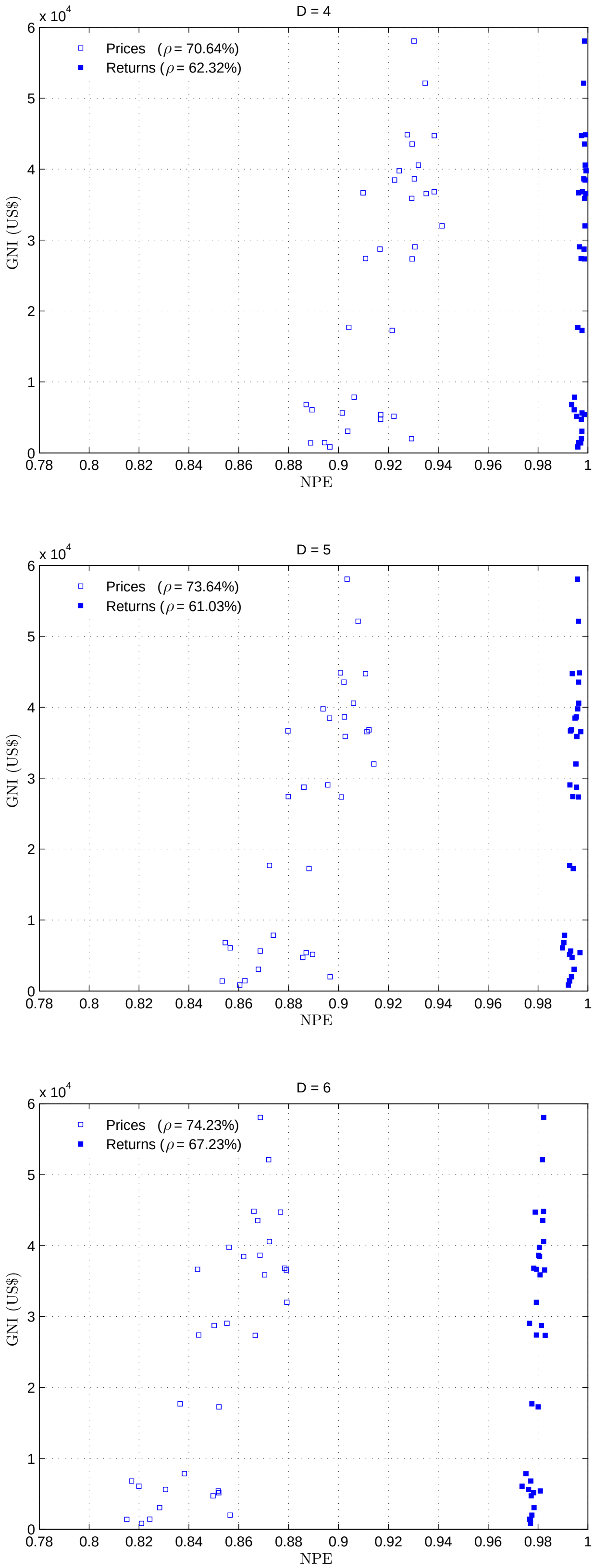

Fig. 3. The Gross National Income (GNI) per capita as a function of the NPE for prices and returns with embedding dimension $D = 4$ (top), $D = 5$ (center) and $D = 6$ (bottom).