



A LETTERS JOURNAL EXPLORING
THE FRONTIERS OF PHYSICS

OFFPRINT

Symbolic dynamics techniques for complex systems: Application to share price dynamics

DAN XU and CHRISTIAN BECK

EPL, **118** (2017) 30001

Please visit the website
www.epljournal.org

Note that the author(s) has the following rights:

- immediately after publication, to use all or part of the article without revision or modification, **including the EPLA-formatted version**, for personal compilations and use only;
- no sooner than 12 months from the date of first publication, to include the accepted manuscript (all or part), **but not the EPLA-formatted version**, on institute repositories or third-party websites provided a link to the online EPL abstract or EPL homepage is included.

For complete copyright details see: <https://authors.eplletters.net/documents/copyright.pdf>.



epl

A LETTERS JOURNAL EXPLORING
THE FRONTIERS OF PHYSICS

AN INVITATION TO SUBMIT YOUR WORK

epljournal.org

The Editorial Board invites you to submit your letters to EPL

EPL is a leading international journal publishing original, innovative Letters in all areas of physics, ranging from condensed matter topics and interdisciplinary research to astrophysics, geophysics, plasma and fusion sciences, including those with application potential.

The high profile of the journal combined with the excellent scientific quality of the articles ensures that EPL is an essential resource for its worldwide audience. EPL offers authors global visibility and a great opportunity to share their work with others across the whole of the physics community.

Run by active scientists, for scientists

EPL is reviewed by scientists for scientists, to serve and support the international scientific community. The Editorial Board is a team of active research scientists with an expert understanding of the needs of both authors and researchers.



epljournal.org

OVER

568,000

full text downloads in 2015

18 DAYS

average accept to online
publication in 2015

20,300

citations in 2015

*"We greatly appreciate
the efficient, professional
and rapid processing of
our paper by your team."*

Cong Lin
Shanghai University

Six good reasons to publish with EPL

We want to work with you to gain recognition for your research through worldwide visibility and high citations. As an EPL author, you will benefit from:

- 1 Quality** – The 60+ Co-editors, who are experts in their field, oversee the entire peer-review process, from selection of the referees to making all final acceptance decisions.
- 2 Convenience** – Easy to access compilations of recent articles in specific narrow fields available on the website.
- 3 Speed of processing** – We aim to provide you with a quick and efficient service; the median time from submission to online publication is under 100 days.
- 4 High visibility** – Strong promotion and visibility through material available at over 300 events annually, distributed via e-mail, and targeted mailshot newsletters.
- 5 International reach** – Over 3200 institutions have access to EPL, enabling your work to be read by your peers in 100 countries.
- 6 Open access** – Articles are offered open access for a one-off author payment; green open access on all others with a 12-month embargo.

Details on preparing, submitting and tracking the progress of your manuscript from submission to acceptance are available on the EPL submission website epletters.net.

If you would like further information about our author service or EPL in general, please visit epijournal.org or e-mail us at info@epijournal.org.

EPL is published in partnership with:



European Physical Society



Società Italiana
di Fisica



EDP Sciences



IOP Publishing

Perspective

Symbolic dynamics techniques for complex systems: Application to share price dynamics

DAN XU and CHRISTIAN BECK

Queen Mary University of London, School of Mathematical Sciences - Mile End Road, London E1 4NS, UK

received 15 June 2017; accepted in final form 20 June 2017

published online 7 July 2017

PACS 05.45.-a – Nonlinear dynamics and chaos

PACS 05.45.Tp – Time series analysis

PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion

Abstract – The symbolic dynamics technique is well known for low-dimensional dynamical systems and chaotic maps, and lies at the roots of the thermodynamic formalism of dynamical systems. Here we show that this technique can also be successfully applied to time series generated by complex systems of much higher dimensionality. Our main example is the investigation of share price returns in a coarse-grained way. A nontrivial spectrum of Rényi entropies is found. We study how the spectrum depends on the time scale of returns, the sector of stocks considered, as well as the number of symbols used for the symbolic description. Overall our analysis confirms that in the symbol space transition probabilities of observed share price returns depend on the entire history of previous symbols, thus emphasizing the need for a modelling based on non-Markovian stochastic processes. Our method allows for quantitative comparisons of entirely different complex systems, for example the statistics of symbol sequences generated by share price returns using 4 symbols can be compared with that of genomic sequences.

perspective

Copyright © EPLA, 2017

Introduction. – The symbolic dynamics technique is a powerful method to describe trajectories of a dynamical system in a coarse-grained way [1–10]. It has been applied to many low-dimensional maps exhibiting chaotic or critical behavior. Nontrivial correlations in the dynamical system manifest themselves in a nontrivial spectrum of Rényi entropies [1,4,5,11–13] and other observables associated with the set of allowed symbol sequences and their corresponding probabilities. This technique, which was very popular in the 1980s and 1990s when a lot of research on 1-dimensional maps was done [2–5,10], has been revived in more recent work and applied in a more general context [14–17].

Our main aim in this paper is to illustrate that the symbolic dynamics technique borrowed from dynamical systems theory can be successfully applied to time series generated by general complex systems in higher dimensions, far beyond the original 1-dimensional chaotic map approach. Our main example in the following will be share price dynamics, which is of course ultimately produced by a complex market and trader dynamics in a high-dimensional phase space [18]. It is well known that financial time series exhibit multifractal features [19–22],

but here we present a somewhat different approach to this problem based on the symbolic dynamics technique. We will investigate the correlations and complex behavior associated with discrete symbol sequences generated from observed share price returns on various time scales. We will quantify this by the calculation of the corresponding Rényi entropies [1,11–13] for symbol sequences based on historical data sets of share price returns for various sectors, both on long (daily) time scales and on short time scales (minutes). It turns out that the stochastic process of symbol sequences observed for real share price data exhibits non-Markovian character. To characterize differences between various companies (or communities in a complex system context), we will introduce a Rényi difference matrix which compares Rényi entropies of different subsystems. The method developed allows for quantitative comparisons of different complex subsystems, or even different scientific problems due to the encoding in the symbolic dynamics space. For example, it is possible to quantitatively compare the statistical properties of genomic sequences [23–25] with those generated by coarse-grained share price movements, although both problems come from completely different areas of science.

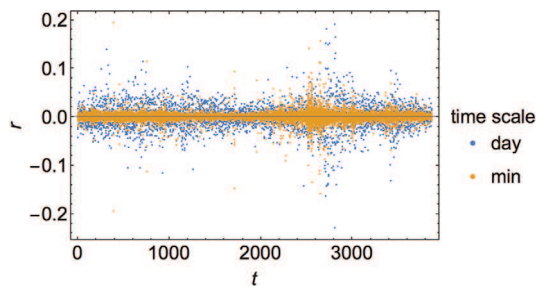


Fig. 1: (Colour online) Time series of log returns of a share price (in this example Alcoa Inc.) over the period from 1998 to 2013, exhibiting intermittent outbursts of strong volatility. The time unit of t consists in trading days. A year has about 251 trading days.

Symbolic dynamics for share prices. – Let us apply the symbolic dynamics technique, well known from dynamical systems theory, to a coarse-grained description of a given time series. This time series is assumed to be generated by a suitable observable of a dynamically evolving complex system. Rather than being very theoretical, we choose a concrete example: Share price evolution, as created by complex market structures and trader speculations. Often, one is interested only in a very basic question related to a given complex system. For the example of share prices, this question is quite straightforward: Basically one is interested in whether the share price of a particular company will go up or down, and this problem of course also depends on the time scale considered. To generate symbol sequences associated with a question of this type we first need to choose a suitable partition of the phase space (a generating partition in the case of low-dimensional maps).

The easiest and most straightforward way for share prices is to consider just two symbols, *i.e.*, to consider the possible values of logarithmic share price returns $\log \frac{S_{n+1}}{S_n}$ in two disjoint subsets $A_1 = (-\infty, 0)$ and $A_2 = (0, \infty)$, corresponding to negative and positive changes of the share price S_n on a discrete time scale labelled as n . This is kind of a reduced phase-space description for the question asked here. Note that A_1 and A_2 are chosen as open intervals. The point 0 (corresponding to no change) has measure 0, it does not influence the analysis. Of course, for other complex systems/time series one can define the subsets generating the symbol sequences differently, which depends on the problem and the question asked about the complex system. In general, much more complicated generating partitions arise in this way [9,14–17].

Let us now look at an example data set of daily stock prices S_n of the company Alcoa Inc. that covers the period from January 1998 to May 2013 (fig. 1). If the log return is an element of A_1 , which is equivalent to $S_{n+1} < S_n$, then we denote such a price-decrease event by the symbol d . Otherwise a price in A_2 stands for $S_{n+1} > S_n$ and is denoted by u which means a price increase. By this method we can attribute to the time series of share prices

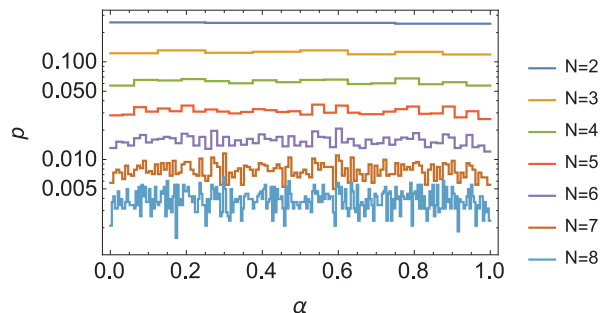


Fig. 2: (Colour online) Joint probabilities $p_j^{(N)}$ of the daily share prices movement dynamics for symbol sequence of length from 2 to 8 for Alcoa Inc.

a symbol sequence $i_0, i_1, i_2, \dots, i_n, \dots$, where $i_n \in \{d, u\}$. We now consider subsequences of symbols of length N , where N is small as compared to the total number of data available. As we only have data for a limited set of data points, we divide up the whole data sequence into R segments of equal length N . Because each symbol has only two choices which are either u or d , for any given N we get up to $\omega(N) = 2^N$ allowed subsequences. Since the partition of the phase space is rather simple and our dataset is big enough to satisfy $R \gg \omega(N)$, there will be many occurrences that correspond to the same symbolic pattern. Hence we can easily acquire the probabilities of each allowed symbol sequence by determining the frequencies of how often the symbol sequence occurs in the given data set. The probability of a given symbol sequence of length N is then denoted as $p_j^{(N)} = p(i_0, \dots, i_{N-1})$, where j labels all possible sequences i_0, \dots, i_{N-1} .

Now by encoding each allowed symbol sequence of length N into a real number α on the unit interval using the binary expansion with $0 = d$ and $1 = u$, we can produce a plot to visualise our probabilities. This means that any given sequence of symbols i_0, \dots, i_{N-1} can be represented by a sequence of bits, in particular we assign

$$x_{i_n} = \begin{cases} 1, & \text{if } i_n = u, \\ 0, & \text{if } i_n = d, \end{cases} \quad (1)$$

where $n = 0, 1, \dots, N-1$. One can implement this by defining a coordinate α assigned to a given symbol sequence as

$$\alpha(\mathbf{x}^{(N)}) = \sum_{n=1}^N x_{i_{n-1}} 2^{-n}. \quad (2)$$

Note that $\alpha(\mathbf{x}^{(N)}) \in [0, 1)$. In this way we allocate to each symbol sequence a real number α on the unit interval so that we can easily visualize our results on the frequency of a given symbol sequence. This is shown in fig. 2.

Note that numerically we restrict ourselves to $N \leq 8$ because of the fact that we only have a limited amount of data points. Larger N would induce large stochastic errors due to the fact that the statistics is not high enough to

estimate the frequency of a given symbol sequence in a reliable way.

Also note that while for the simple example considered here all symbol sequences are possible, for more general complex systems (and more complicated questions asked that are then encoded into the symbols, such as the problems in [14–17]) there will be a set of allowed and forbidden sequences in the symbol space. In this case the distribution in fig. 2 will have gaps and generally there will be a multifractal structure. Still everything that we define in this paper can still be done in an analogous way.

Information contents of the symbolic sequences.

– Given experimentally determined probability distributions with possible self-similar features such as in fig. 2, it is meaningful to find a proper way to measure multifractal features. For this purpose we use the well-known concept of *Rényi information* [1,11–13] defined as

$$I_q(p^{(N)}) = \frac{1}{q-1} \ln \sum_{j=1}^{\omega(N)} (p_j^{(N)})^q. \quad (3)$$

Here q is a real number and $\omega(N)$ is the number of allowed symbol sequences for a given N (an allowed symbol sequence is one that satisfies $p_j^{(N)} \neq 0$). The index j labels the various sequence probabilities. The Rényi information measure can be regarded as a generalisation of the Shannon information, as for $q \rightarrow 1$ we have

$$\lim_{q \rightarrow 1} I_q(p^{(N)}) = \sum_{j=1}^{\omega(N)} p_j^{(N)} \ln p_j^{(N)} = I(p^{(N)}), \quad (4)$$

where $I(p^{(N)})$ denotes the Shannon information.

An important special case in eq. (3) is the choice $q = 0$,

$$I_0(p^{(N)}) = -\ln \omega(N), \quad (5)$$

which means $I_0(p^{(N)})$ decreases in a logarithmical way with the number of allowed symbol sequences. Note that when we have a limited length N of symbol sequences, eq. (3) yields a finite value as $\omega(N)$ is finite.

Recall that for a given N any symbol sequence is mapped onto a point of $[0, 1)$, with equal distance between neighboring points. We call the distance of two neighboring coordinates at a given level in fig. 2 *box size* and denote it by ε . In our case the box size is determined by N as $\varepsilon = \frac{1}{2^N}$. This means that ε is getting smaller as N grows and when $N \rightarrow \infty$, ε is approaching 0, in which case the Rényi information defined by eq. (3) diverges. It is then useful to define the *Rényi dimension* which is a useful quantity as it stays finite in the limit $\varepsilon \rightarrow 0$:

$$D(q) = \lim_{\varepsilon \rightarrow 0} \frac{I_q(p^{(N)})}{\ln \varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\ln \varepsilon} \frac{1}{q-1} \ln \sum_{j=1}^{\omega(N)} (p_j^{(N)})^q. \quad (6)$$

An example of (finite box size) Rényi dimensions evaluated for the data given in fig. 2 is shown in fig. 3.

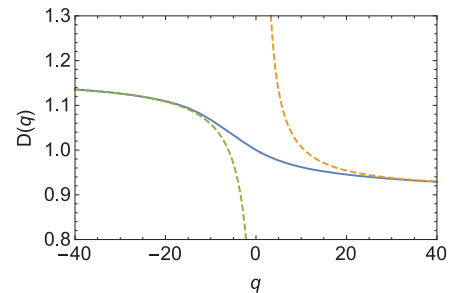


Fig. 3: (Colour online) Rényi dimensions together with the upper (orange dashed line) and lower (green dashed line) bounds for daily share prices movement of shares of Alcoa Inc (as numerically obtained for $N = 8$).

Of course the size of any data set is limited, in our case the smallest ε that can be achieved with reliable non-fluctuating results is $\frac{1}{2^8}$. Given the finiteness of the data set it is useful to check some rigorous upper and lower bounds and monotonicity properties of the Rényi dimensions, valid for arbitrary probability measures. The Rényi dimensions are monotonically decreasing and their values must be positive for all q . Also, if all symbol sequences are allowed then we must get the value 1 when $q = 0$. This is because

$$D(0) = \lim_{\varepsilon \rightarrow 0} \left(-\frac{1}{\ln \varepsilon} \right) \ln \sum_{j=1}^{\omega(N)} 1 = \lim_{\varepsilon \rightarrow 0} \left(-\frac{\ln 2^N}{\ln \frac{1}{2^N}} \right) = 1. \quad (7)$$

In addition, as shown in [12] there is a restriction of possible values of the Rényi dimensions as a general upper and lower bound can be proved:

$$\frac{q'-1}{q'} D(q') \geq \frac{q-1}{q} D(q), \quad \text{for } q' > q, \quad q'q > 0. \quad (8)$$

If we substitute $+\infty$ and $-\infty$ for q in eq. (8), we obtain the upper bound

$$D(q) \leq \frac{q}{q-1} D(+\infty), \quad \text{for } q > 1, \quad (9)$$

and a lower bound is given by

$$D(q) \geq \frac{q}{q-1} D(-\infty), \quad \text{for } q < 0. \quad (10)$$

We have checked these bounds for our data set with $N = 8$, the result is also shown in fig. 3. Clearly the inequalities (9) and (10) are satisfied by our experimental data.

From the Rényi dimensions $D(q)$ one can proceed to the singularity spectrum $f(\tilde{\alpha})$ of crowding indices $\tilde{\alpha}$ by a Legendre transformation in the thermodynamic formalism. However, as the information contained in $f(\tilde{\alpha})$ spectra is the same as the one in the generalized dimensions $D(q)$, we will not further proceed along these lines here, but refer the reader to suitable literature introducing to this topic [1,26].

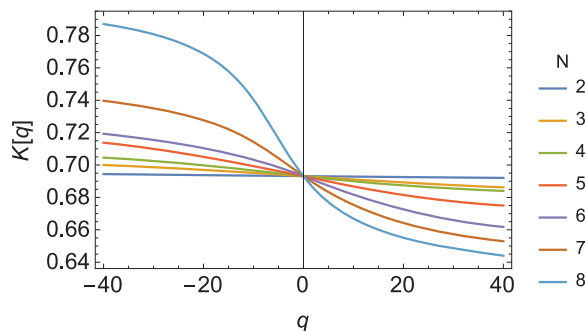


Fig. 4: (Colour online) Rényi entropies for the daily share price movement dynamics for symbol sequence of length from 2 to 8 for Alcoa Inc.

Rényi entropies. – Unlike the Rényi dimension, which generally is a property of a given multifractal probability measure and which does not *a priori* contain any dynamical information, the *Rényi entropy* $K(q)$ measures the production (or loss) of information encoded in the symbol sequences. This is the quantity of direct interest for share price evolution. The Rényi entropy $K(q)$ is defined in the limit $N \rightarrow \infty$ and given by

$$K(q) = \lim_{N \rightarrow \infty} \frac{-I_q^{(N)}}{N} = \lim_{N \rightarrow \infty} \frac{1}{1-q} \frac{1}{N} \ln \sum_{j=1}^{\omega(N)} (p_j^{(N)})^q. \quad (11)$$

Of course, by projecting symbol sequences onto points on the real line (as done in fig. 2), both approaches can be made formally equivalent, but the dynamical information is then just encoded in a suitable multifractal measure.

Figure 4 shows finite- N versions of Rényi entropies for share price returns where N grows from 2 to 8. The Rényi entropies have of course the same dependence on q as the Rényi dimensions of the corresponding multifractal measure that encodes the symbol sequence probabilities on the unit interval.

For $q = 0$ we get

$$K(0) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln 2^N = \ln 2, \quad (12)$$

which is the *topological entropy*. Moreover, for $q \rightarrow 1$, using eq. (4),

$$K(1) = \lim_{N \rightarrow \infty} \frac{-I(p^{(N)})}{N}, \quad (13)$$

we get the *Kolmogorov-Sinai entropy* associated with the symbol sequences of share price changes. Again one can proceed from the function $K(q)$ to an equivalent spectrum of dynamical indices $g(\tilde{\alpha})$ by Legendre transformation (see, e.g., [1]).

Small time scales. – We have quantified the joint probabilities of the daily share price movements by the corresponding Rényi entropies. We are now interested in how

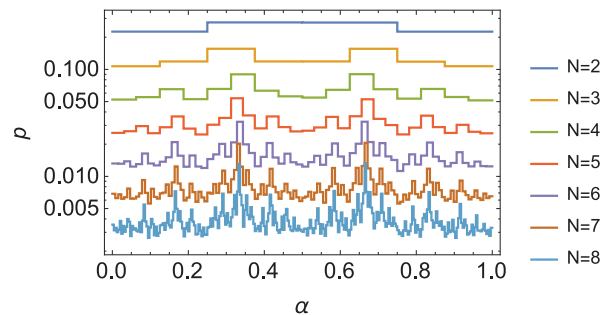


Fig. 5: (Colour online) Joint probabilities of share price dynamics for symbol sequence of length from 2 to 8 for Alcoa Inc., evaluated on a time scale of minutes.

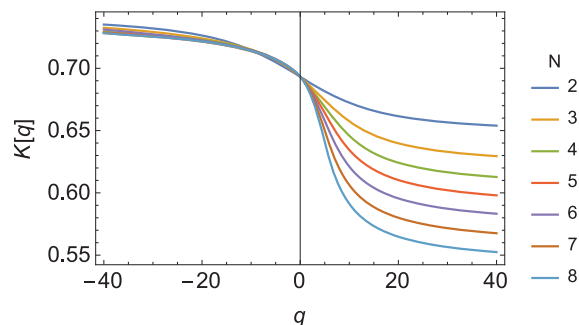


Fig. 6: (Colour online) The Rényi entropies as obtained for Alcoa shares on a time scale of minutes, with $N = 2, \dots, 8$.

these observables depend on the time scale of the symbolic dynamics. Instead of using the daily share prices, the now analysed data set consists of share prices recorded each minute for the same period from 1998 to 2013; this covers about 1.5 million data points. We repeat the same analysis as before and consider symbol sequences up to length 8. By using the same partition method as in the previous section we obtain the multifractal probability distributions of symbol sequences as shown in fig. 5.

Compared with fig. 2, the probability distributions on a small time scale are significantly different from those on the daily time scale. Note that there are some local maxima which are reproduced in a self-similar way. While these densities are non-smooth, the advantage of proceeding to the Rényi dimensions (or entropies) is that in this way a smooth dependence on the scanning parameter q is produced. This is shown in figs. 6 and 7.

We see, similarly to the daily time scale, that both the Rényi entropies and the Rényi dimensions are monotonically decreasing with respect to the parameter q , with larger N generating a more pronounced q -dependence for positive q , whereas there is hardly any N -dependence for $q < 0$. Figure 7 also shows the upper and lower bound.

An important property that we have checked for our data sets is the fact that quite generally the conditional probabilities

$$p(i_N | i_0, \dots, i_{N-1}) = p(i_0, \dots, i_N) / p(i_0, \dots, i_{N-1}) \quad (14)$$

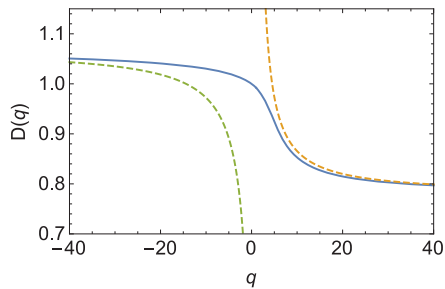


Fig. 7: (Colour online) Upper and lower bounds (dashed lines) for the Rényi dimensions (solid line) of Alcoa shares on a time scale of minutes ($N = 8$).

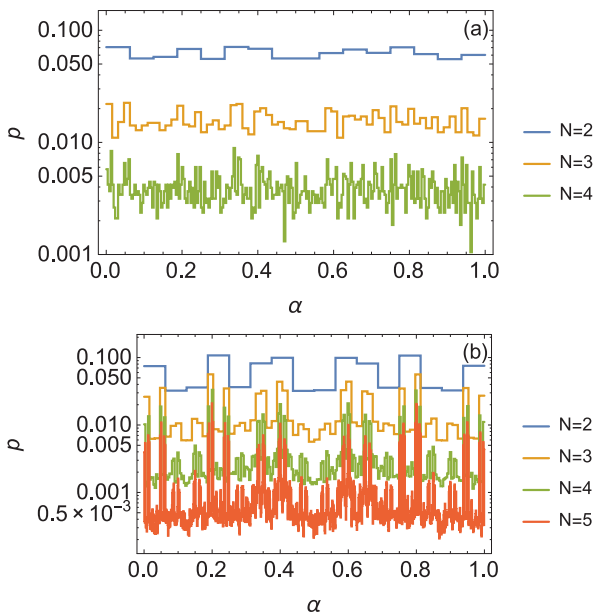


Fig. 8: (Colour online) Symbol sequence probabilities of length N for Alcoa shares using an alphabet of 4 symbols (daily (a) and minute (b) time scale).

depend on the entire history i_0, \dots, i_{N-1} , *i.e.*, a Markovian model cannot capture the complex features in the symbol space. For example, given a long alternating sequence u, d, u, d it is statistically slightly more likely to observe the next symbol as u .

4-symbol partitions. – So far we used the simplest method to study the symbolic dynamics of the share price movements, just considering whether a share price goes up or down. We may, however, also ask a more detailed question, such as whether the share price goes up slightly or strongly. To detect further details, we may generate a refined version of the phase-space partition $\mathbf{A} = \{A_1, A_2\}$ where previously $A_1 = (-\infty, 0)$ and $A_2 = (0, \infty)$. Assume there exists a real number c where the log returns have equal 1-point probabilities to fall into each element of a partition $\mathbf{B} = \{B_1, B_2, B_3, B_4\}$ given by

$$B_1 = (-\infty, -c), B_2 = (-c, 0), B_3 = (0, c), B_4 = (c, +\infty). \quad (15)$$

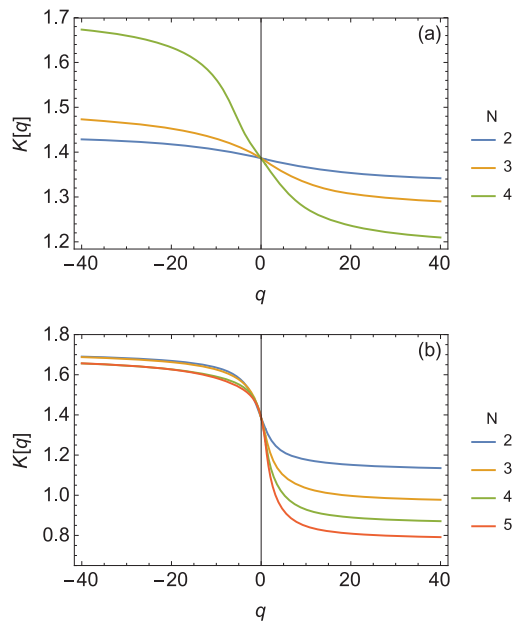


Fig. 9: (Colour online) Finite- N Rényi entropies for the same data as in fig. 8(a), (b).

In other words, this partition is chosen in such a way that the 1-point probabilities $p(B_i)$, $i = 1, 2, 3, 4$ of the log returns lying in each set B_i are identical to $1/4$. For Alcoa Inc. share prices on daily and minutely time scales, we find c is equal to 0.014 and 0.00088, respectively. Instead of denoting the time evolution by u and d , we redefine our symbol sequence by

$$i_0, i_1, \dots, i_n, \dots, \quad i_n \in \{b_1, b_2, b_3, b_4\}, \quad (16)$$

where b_i corresponds to a log return in B_i . For a given length N , the number of allowed sequences is $\omega(N) = 4^N$. We may also upgrade the previous approach to a 4-level symbolic sequence given by

$$x_{i_n} = \begin{cases} 0, & \text{if } i_n = b_1, \\ 1, & \text{if } i_n = b_2, \\ 2, & \text{if } i_n = b_3, \\ 3, & \text{if } i_n = b_4. \end{cases} \quad (17)$$

A symbol sequence of length N can be encoded as a coordinate α on the unit interval based on the representation

$$\alpha(\mathbf{x}^{(N)}) = \sum_{n=1}^N x_{i_{n-1}} 4^{-n}, \quad \alpha(\mathbf{x}^{(N)}) \in [0, 1]. \quad (18)$$

The plot of joint probabilities in the case of an alphabet of 4 symbols is shown in fig. 8, the corresponding finite- N Rényi entropies are shown in fig. 9.

Different companies. – So far we only looked at a particular example of a stock, Alcoa shares. The really interesting work on complex system analysis starts when one compares Rényi entropies of different companies, or

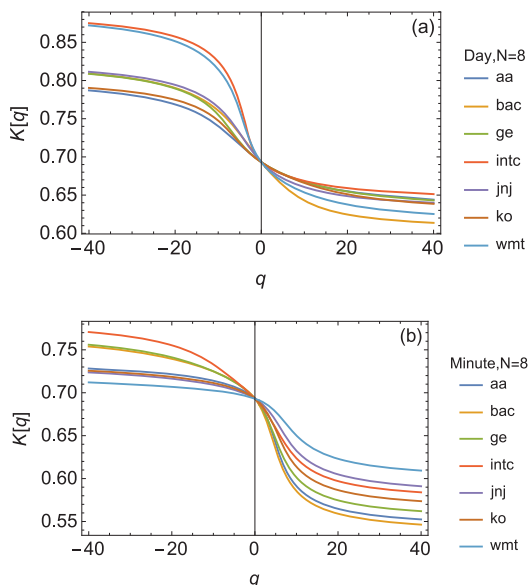


Fig. 10: (Colour online) Rényi entropies for 7 different companies on daily scale (a) and minute scale (b). The alphabet contains 2 symbols.

even Rényi entropies of entirely different complex systems described by the same alphabet of symbols. Ultimately we really want to learn something about the complex market structure and dynamics represented by different companies, sectors, or communities for general complex systems. Still one can let each community generate a suitable time series for a suitable observable and then analyse this time series with the symbolic dynamics technique described so far. We are then interested in differences or similarities of the obtained Rényi entropies.

To illustrate this we looked at symbol sequences as generated by share prices of 7 different companies, Alcoa (aa), Bank of America (bac), General Electric (ge), Intel (intc), Johnson & Johnson (jnj), Coca Cola (ko), WalMart (wmt), representing the sectors basic materials, financial, industrial goods, technology, healthcare, consumer goods, services. The results are shown in fig. 10. It can be observed that bac appears to have the lowest Rényi entropy in the region $q > 0$ as compared to other stocks, overall the q -dependence for financial stocks is most pronounced. This could have to do with the fact that financial stocks have relatively strong fluctuations and exhibit nontrivial correlations, described by a non-trivial spectrum of Rényi entropies. In any case, different companies are characterized by a different spectrum of Rényi entropies both on a daily (fig. 10(a)) and minute (fig. 10(b)) time scale. The smooth dependence on the parameter q can be used for an effective thermodynamic description of the complex behavior involved, with different emphasis given to low and high probabilities depending on the value of the scanning parameter q .

Quantifying similarities in the symbol sequence statistics. – We may now wish to compare in a

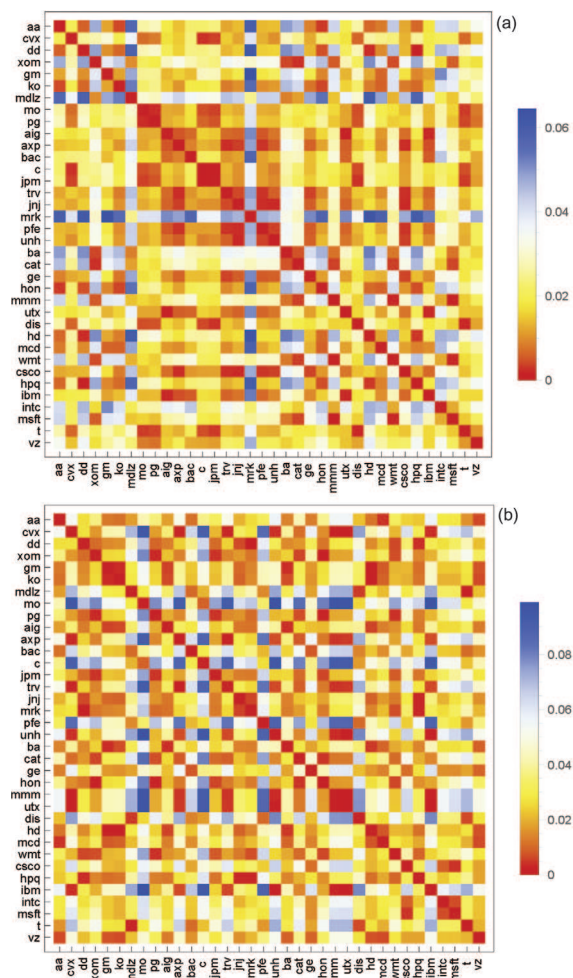


Fig. 11: (Colour online) Rényi difference matrix R_{ij} on a daily time scale (a) and minute time scale (b) for 36 stocks traded at the NYSE-Nasdaq. The value of the parameter κ was chosen as $\kappa = 1$. Similar pictures arise for other values of κ .

quantitative way how much the Rényi entropies of different companies (or communities in the general complex system context) differ. For this purpose we define a *Rényi difference matrix* R_{ij} as follows:

$$R_{ij} = \frac{1}{q_{max} - q_{min}} \int_{q_{min}}^{q_{max}} |K_i(q) - K_j(q)|^\kappa dq. \quad (19)$$

Clearly, if two companies i and j have the same statistics of symbol sequences, described by the same functional dependence $K_i(q) = K_j(q)$, then the Rényi difference matrix element is $R_{ij} = 0$. Otherwise, the entry R_{ij} integrates up differences in the Rényi entropy spectra, and averages them over q , weighted with the parameter κ .

Figure 11 shows a colour encoding of such a Rényi difference matrix. For 36 different companies we evaluated R_{ij} , choosing $\kappa = 1$ and $q_{min} = -40, q_{max} = +40$. The Rényi difference matrix allows one to single out major differences and similarities in the symbol sequence statistics of different companies/communities in a quantitative way.

In our case, for example, the healthcare company Merck (mrk) is identified as having an unusual Rényi entropy spectrum on a daily scale, different from that of most other companies, visible here as a pronounced vertical and horizontal blue line in the pattern generated. On the other hand, on the small time scale of minutes this company is much more similar to the others.

An interesting final remark is at order. Once a suitable question has been asked about a complex system, and a symbolic dynamics constructed, we can then compare different complex systems, or subsystems thereof, whatever their origin, as all relevant information is encoded in the symbol sequence statistics. To give an example, in [24,25] the complexity of DNA sequences of the human genome was investigated, by calculating the Rényi entropies associated with the sequence statistics of the 4 bases A, T, G, C. Once this function is obtained, it can then be compared using the above Rényi difference matrix with other genomes, in just the same way as we compared the differences between different companies in fig. 11. But more drastically, we can even compare in a quantitative way (via R_{ij}) the Rényi spectra of completely different complex systems, such as the complexity of financial markets and the complexity of genomes. This is the advantage of the symbolic dynamics encoding technique: Once a function $K_i(q)$ has been obtained, one can compare it in a quantitative way with another function $K_j(q)$, whatever its origin, and thus measure differences in complexity and information production in a quantitative way.

Conclusions and outlook. – Although the examples considered in this paper were all based on symbol sequences generated by share price returns, it is clear that the same method can be applied to symbol sequences generated by time series of all kinds of complex systems, whatever their origin. In this way the Rényi entropies associated with such a symbolic description allow for a quantitative comparison of the dynamical properties in the symbol space, making it easy to compare different complex systems, or different substructures/communities within a given big complex system. In fact, in this way one can compare entirely different complex systems, for example the Rényi entropies associated with share price changes (using an alphabet of 4 symbols) can be compared with those of genomic sequences [24,25] or those of successive quantum-mechanical measurements [27]. The most important dynamical information is then encoded in the form of the shape of the function $K(q)$, allowing the application of thermodynamic tools [1]. In this way a quantitative comparison of different systems is possible, solely based on

the Rényi information contents of the coarse-grained symbolic description. The extension of the methods presented here to more complicated symbolic dynamics generated by other types of complex systems is straightforward.

REFERENCES

- [1] BECK C. and SCHLÖGL F., *Thermodynamics of Chaotic Systems* (Cambridge University Press) 1993.
- [2] COLLET P. and ECKMANN J.-P., *Iterated Maps on the Interval as Dynamical Systems* (Birkhauser) 1980.
- [3] HAO BAI-LIN, *Elementary Symbolic Dynamics and Chaos in Dissipative Systems* (World Scientific) 1989.
- [4] TÉL T., *Z. Naturforsch.*, **43a** (1988) 1154.
- [5] CSORDÁS P. and SZÉPFALUSY P., *Phys. Rev. A*, **39** (1989) 4767.
- [6] LIND D. and MARCUS B., *An Introduction to Symbolic Dynamics and Coding* (Cambridge University Press) 1995.
- [7] CRUTCHFIELD J. P. and PACKARD N. H., *Int. J. Theor. Phys.*, **21** (1982) 433.
- [8] RUELLE D., *Bol. Soc. Bras. Mat.*, **9** (1978) 83.
- [9] GRASSBERGER P. and KANTZ H., *Phys. Lett. A*, **113** (1985) 235.
- [10] FEIGENBAUM M. J., *J. Stat. Phys.*, **19** (1978) 25; **21** (1979) 669.
- [11] RÉNYI A., *Probability Theory* (North-Holland Amsterdam) 1970.
- [12] BECK C., *Physica D*, **41** (1990) 67.
- [13] GODÓ B. and NAGY Á., *Chaos*, **26** (2016) 083102.
- [14] POLITI A., *Phys. Rev. Lett.*, **118** (2017) 144101.
- [15] LACASA L., NICOSIA V. and LATORA V., *Sci. Rep.*, **5** (2015) 15508.
- [16] LACASA L. and JUST W., arXiv:1704.06467.
- [17] SMITH S. A., ARENSON J., ROBERTS E., SINDI S. and MITCHELL K. A., *EPL*, **117** (2017) 60005.
- [18] BIONDO A. E., PLUCHINO A., RAPISARDA A. and HELBING D., *Phys. Rev. E*, **88** (2013) 062814.
- [19] MORALES R., DI MATTEO T. and ASTE T., *Sci. Rep.*, **4** (2014) 4589.
- [20] BACRY E., DELOUR J. and MUZZY J. F., *Physica A*, **299** (2001) 84.
- [21] KANTELHARDT W. *et al.*, *Physica A*, **316** (2002) 7.
- [22] JIZBA P. and KORBEL J., *Int. J. Des. Nat. Ecodyn.*, **10** (2015) 261 (arXiv:1610.07028).
- [23] YU Z. G., ANH V. and LAU K. S., *Phys. Rev. E*, **64** (2001) 031903; **68** (2003) 021913.
- [24] BECK C. and PROVATA A., *EPL*, **95** (2011) 58002.
- [25] PROVATA A. and BECK C., *Phys. Rev. E*, **83** (2011) 066210.
- [26] HALSEY T. C., JENSEN M. H., KADANOFF L. P., PROCACCIA I. and SHRAIMAN B. I., *Phys. Rev. A*, **33** (1986) 1141.
- [27] BECK C. and GRAUDENZ D., *Phys. Rev. A*, **46** (1992) 6265.