Recurrence Plot (RP) and Recurrence Quantification Analysis (RQA) are signal numerical analysis methodologies able to work with nonlinear dynamical systems and nonstationarity. Moreover, they well evidence changes in the states of a dynamical system. We recall their features and give practical recipes. It is shown that RP and RQA detect the critical regime in financial indices (in analogy with phase transition) before a bubble bursts, whence allowing to estimate the bubble initial time. The analysis is made on DAX and NASDAQ daily closing price between January 1998 and November 2003. DAX is studied in order to set-up overall considerations, and as a support for deducing technical rules. The NASDAQ bubble initial time has been estimated to be on 19 October 1999.

Keywords: Endogenous crash; financial bubble; Recurrence Plot; Recurrence Quantification Analysis; nonlinear time series analysis; DAX; NASDAQ.

1. Introduction

The occurrence of a financial crash is a tragic event for many investors. Recent papers have shown some analogy between crashes and phase transitions.\textsuperscript{1–3} Moreover, like in earthquakes, log periodic oscillations have been found before some crashes.\textsuperscript{4,5} It was then proposed that an economic index $y(t)$ increases as a complex power law, whose first order Fourier representation is

$$y(t) = A + B \ln(t_c - t) \{1 + C \cos[\omega \ln(t_c - t) + \phi]\},$$

where $A$, $B$, $C$, $\omega$, $\phi$ are constants and $t_c$ is the critical time (rupture time). From best fits with this formula, it has been shown that one can predict financial crashes.\textsuperscript{4} Notwithstanding, this does not answer the question why a crash occurs. A crash is thought to occur when the system goes out of equilibrium and enters an unstable phase where any information is amplified.\textsuperscript{6} It should be noted that a crash
necessarily happens during a period of generalized economic euphoria, with high returns, high volatility, high correlation between markets and unprecedent imitation behavior by investors. This critical period takes the name of “speculative bubble”.

A financial time series is often studied from the point of view of stochastic processes. On the other hand, it is well known that nonlinear simple deterministic systems can exhibit the phenomenon of deterministic chaos. The debate on the appropriate approach, i.e., if it has to be stochastic or instead must see these events from the point of view of deterministic chaos, generated by nonlinear dynamics, is quite open. A topological approach to analyze chaos developed in the physics literature based on close return plots, for example, has produced much evidence that financial data are not chaotic, where chaotic is meant as something unordered, unpredictable and confusing.

We have investigated the possibility to use Recurrence Plot (RP) and Recurrence Quantification Analysis (RQA) techniques for financial time series, in particular for detecting critical regimes preceding an endogenous crash. Recurrence Plots are graphical tools elaborated by Eckmann, Kamphorst and Ruelle in 1987 and are based on Phase Space Reconstruction. In 1992, Zbilut and Webber proposed a statistical quantification of RPs and gave it the name of “Recurrence Quantification Analysis” (RQA). Here RP and RQA techniques are not intended to prove evidence of chaos, but are used for their goodness in working with nonstationarity and noisy data, in detecting changes in data behavior, in particular in detecting breaks, like a phase transition, and in informing about other dynamic properties of a time series.

Most of the applications of RP and RQA are at this time in the field of physiology and biology as for example, protein folding analysis, but some authors have already applied the techniques to financial data and cloud base height fluctuation. In financial application, some authors have chosen these techniques for their goodness to work with nonlinearity or they applied to foreign exchange data for no stationarity, others for testing chaos or noise. Thiel et al. in Ref. 30 have tested the quantity of information contained in RP claiming that it is possible to reconstruct a time series from its unthresholded RP. Even if the choice of threshold and embedding parameters can condition the quantity of information that can be read in an RP, “... this new framework provides a useful complementary tool to test for complexity in financial data ...” A complete bibliography on RP and RQA and their application can be found in the web page by Marwan, while a recent overview is found in Ref. 32.

The following work is an application of RP and RQA techniques in studying rare events, like crashes in financial markets and in particular to search whether these techniques can distinguish between normal and critical regime, in the thermodynamic sense, whence give an estimation of the initial bubble time. The analysis is here below made on two time series, NASDAQ and DAX, taken over a time span of six years including the known crash of April 2000. The series are also divided into subseries in order to investigate changes in the evolution of the signal.
The treatment is divided into two parts, a theoretical and a practical one. The first part is dedicated to recall the theoretical basis of “Phase Space Reconstruction” (Sec. 2) with attention on the methods used to choose the so-called embedding parameters, to explain how to plot an RP (Sec. 3) and to define the RQA variables (Sec. 4). The formulas for each variable are found in Appendix A. Several examples of RPs and RQA on how to interpret the results are also found and discussed in Secs. 3 and 4. Moreover a simulation of an economophysically interesting log-periodic signal generated from Eq. (1) has been made and is discussed in Sec. 4.

The second part is dedicated to our specific analysis. In Sec. 5, the data and the methodology analysis are presented. In Sec. 6, “practical considerations” are outlined about the choice of embedding parameters for financial signals. Analysis of DAX and NASDAQ series are discussed respectively in Secs. 7 and 8. In Sec. 9, conclusions are found.

2. Phase Space Reconstruction

Recurrence Plots are graphical tools based on Phase Space Reconstruction. The changing state of a dynamic system can be indeed represented by sequences of “state vectors” in the phase space, a so called “phase-flow”. Usually in a dynamic system a detailed analysis is possible when the equations of motion and all degrees of freedom \( n \) are known. Unfortunately only a few quantities can be usually observed in a system. However, from a theoretical point of view it is argued that it is possible to reconstruct the entire dynamics of a system from a relatively small number of observables; because the different degrees of freedom of a dynamic system interact with each other, the combination of all other components is concealed in each observable quantity through the main state vector components. Let \( \xi_{t+1} = F(\xi_t) \) be a deterministic system, let \((X_1, \ldots, X_N)\) be \(N\) observations of some variable of this system. According to Takens’ theorem, it is possible to reconstruct a phase space trajectory from such a single scalar time series of observable quantities \((X_1, \ldots, X_N)\). In doing so, it is possible to answer a wide range of questions about the system by examining the dynamics in a space defined by delayed vectors of dimension \(m\). Construct a delayed vector \(y(i) = \{X_i, X_{i+d}, \ldots, X_{i+(m-1)d}\}\), where \(d\) and \(m\) are called respectively the time delay and the embedding dimension, together they are called the embedding parameters. Therefore, each unknown point of the phase space at time \(i\) is reconstructed by the delayed vector \(y(i)\) in a \(m\)-dimensional space called the reconstructed phase space.

2.1. Embedding parameters

The most natural question pertains on how to choose an appropriate value for the time delay \(d\) and the embedding dimension \(m\). Several methods have been developed to best guess \(m\) and \(d\). The most often used methods are the Average Mutual Information function (AMI) for the time delay, as introduced by Fraser...
and Swinney in 1986\textsuperscript{35} and the False Nearest Neighbors (FNN) method for the embedding dimension developed by Kennel et al.\textsuperscript{36}

2.1.1. Time delay

First of all, the time delay has to be estimated, since most logically the method to find the embedding dimension needs an estimation of \( d \).

There are two main methods. In the first one, the value for which the autocorrelation function

\[ C(d) = \frac{1}{N-d} \sum_{i=1}^{N-d} (X_i - \bar{X})(X_{i+d} - \bar{X}) \]

first passes through zero is searched, which gives \( d \). In the second, one chooses the first minimum location of the average mutual information function, where the mutual information function is defined. Let us start with partitioning the real numbers. Let \( p_i \) be the probability to find a time series value in the \( i \)th interval of the partition, let \( p_{ij}(d) \) be the joint probability to find a time series value in the \( i \)th interval and a time series value in the \( j \)th interval after a time \( d \), i.e., the probability of transition in \( d \) time from the \( i \)th to the \( j \)th interval. The average mutual information function is

\[ S(d) = -\sum_{ij} p_{ij}(d) \ln \frac{p_{ij}(d)}{p_i p_j}. \]

The value \( d \) that firstly minimizes the quantity \( S(d) \) is the method choice for finding a reasonable time delay.

The difference between these two methods resides in the fact that while the first looks for linear independence, the second measures a general dependence of two variables. For this reason the second method seems to be preferred in nonlinear time series analysis.

2.1.2. Embedding dimension

The method used to find the embedding dimension is based on the concept of false neighbor. A false neighbor is a point in the data set that looks like a neighbor to another because the orbit is seen in a too small embedding space. For example two points on a circle can appear close to each other, even though they are not, if, e.g., the circle is seen sideways (as a projection), thus is appearing like a line segment, whence, increasing by one the dimension \( m \) of the reconstructed space often permits to differentiate between the points of the orbit, i.e., those which are true neighbors and those which are not.

Let \( y(i) \) be a point of the reconstructed space. Note as \( y(i)^r \) the \( r \)th nearest neighbor and compute the Euclidean distance \( L_2 \) between them.
Table 1. Embedding parameters for simulated data.

<table>
<thead>
<tr>
<th>Embedding parameters</th>
<th>Time delay $d$</th>
<th>emb. dimension $m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brownian motion</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Sine</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>White noise</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Trend</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Log-period</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Function 1</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>

$$R_m^2(y(i), y(i)^r) = \sum_{k=1}^{m-1} [y(i + kd) - y^r(i + kd)]^2.$$

(4)

Next increase $m$ to $m + 1$ and compute the new distance, i.e.,

$$R_{m+1}^2(y(i), y(i)^r) = R_m^2(y(i), y(i)^r) + [y(i + md) - y^r(i + md)]^2.$$

(5)

The point $y^r(i)$ is said to be a false nearest neighbor if

$$\left[ \frac{R_{m+1}^2(y(i), y(i)^r) - R_m^2(y(i), y(i)^r)}{R_m^2(y(i), y(i)^r)} \right] > R_{tol},$$

(6)

where $R_{tol}$ is a predefined threshold. Note that the number of false nearest neighbors depends on $R_{tol}$. The sensitivity of the criterion to $R_{tol}$ is not discussed here. Kennel et al. found that for $R_{tol} \geq 10$, the false nearest neighbor is clearly identified and we stick by this value below, but for a more profound discussion, we suggest to read Ref. 36.

In practice, the percentage of False Nearest Neighbors (FNN) is computed for each $m$ of a set of values; the embedding dimension is said to be found for the first percentage of FNN dropping to zero. Notice that, when the signal is noisy this percentage never reaches a true zero value.

For finding the embedding parameters the TISEAN version 2.1 code by Hegger and Schreiber\textsuperscript{37} was used. TISEAN is a software package for nonlinear time series analysis freely available on the web.\textsuperscript{38}

The values of embedding parameters for the simulated data (Sec. 3.1) are given in Table 1. More practical considerations on the choice of embedding parameters from experimental data with some discussions are left for Sec. 6.

3. Recurrence Plot

The Recurrence Plot (RP) is a matrix of points $(i, j)$ where each point is said to be recurrent and marked with a dot if the distance between the delayed vectors $y(i)$ and $y(j)$ is less than a given threshold $\varepsilon$; the distance can be $L_1$, Euclidean or $L_\infty$.

In general, the threshold $\varepsilon$ has to be chosen as small as possible, but a too small $\varepsilon$ can lead one to miss some structure, if there is noise distortion. As each coordinate $i$ represents a point in time, RP provides information about the temporal correlation of phase space points.\textsuperscript{15} Indeed each horizontal coordinate $i$ in RP refers to the
state of the system at \(i\) and each vertical coordinate \(j\) refers to the state in \(j\). So, if the point \((i, j)\) is marked as recurrent, the state \(j\) belongs to the neighborhood centered in \(i\) of size \(\varepsilon\); this means that the state of the system at time \(i\) has some “similarity” with the state of the system at \(j\), in other words we can say that the system is staying on nearby “orbits”.

The software used for RP and RQA is the Visual Recurrence Analysis (VRA) version 4.6 code by Kononov, freely available on the web.\(^{39}\) We skipped the problem of the threshold using a scale of color for distances; on each plot, white “color” represents a short distance while “black” corresponds to a long distance as it can be seen in the legend on the right side of each plot, in which each “color” indicates a range of distances between states.

### 3.1. Typical cases

For a stationary signal, the dependence should be only on the distance between \(i\) and \(j\) but not on their position, so that an RP should appear homogeneous. Conversely homogeneity in RP indicates stationarity, as seen for example on Fig. 1, showing a Brownian Motion and its RP. By extension, if the texture is homogeneous within an RP block, a stationary behavior can be deduced for the corresponding subperiod.

Vertical or horizontal lines on an RP denote that the system state does not change or change very slowly in time. Diagonal lines correspond to trajectories passing in the same region of the phase space at different times. Therefore, parallel and perpendicular lines to the main diagonal appear when the series presents some determinism or periodicity; see Fig. 2 for a sinusoid and its RP. The length of lines parallel to the main diagonal of the RP indicates how fast the trajectories diverge in phase space.

In pure stochastic systems, no parallel or vertical lines appear on an RP, as can be seen in Figs. 3(b) showing a white noise signal with its RP Fig. 3(a).

In general, an aggregation of points indicates persistent nonstationarity; black bands or areas indicate rare or extreme events.

Therefore RPs can be used to test a system deterministic behavior through the percentage of recurrent points belonging to parallel lines. In doing so, RPs are useful tools for the preprocessing of experimental time series and provide a comprehensive image of the dynamic course at a glance.\(^{18}\) In Appendix A, a summary of typical features can be found.

### 3.2. A log-periodic simulation

In order to study the crash from the point of view of a phase transition with log periodic precursors, i.e., Eq. (1), a log-periodic signal with trend has been simulated. First, the law (1) with \(C = 0\)

\[
y_{\text{trend}}(t) = A + B \ln(t_c - t),
\]

(7)
Fig. 1. (a) The RP of the Brownian Motion for 1000 points plotted in (b). The homogeneity of plot (a) reveals stationarity.
Fig. 2. (a) A sin wave of 1000 points and (b) the corresponding RP. The diagonal lines parallel and vertical to the main diagonal indicating that the signal is periodic. This kind of structure also indicates determinism.
Fig. 3. (a) The RP of a white noise signal, sample of 1000 points, plotted in (b). Plot (a) is homogeneous, without any remarkable pattern; this suggests stationarity and casualty.
is studied; see Fig. 4(b). Then, the contribution of the log-periodicity generated by

$$y_{\text{log-per}}(t) = \cos[\omega \ln(t_c - t) + \phi],$$

(8)
is done; see Fig. 5(b). The simulation of the combination of both terms, i.e., the function (1), is shown in Fig. 6(b).

We can deduce a few interesting points from the displayed features. When a system is not stationary and has a strong trend, an “arrow” shape appears as can be seen in Fig. 4(a), corresponding to the simulation of the function (7). Note the smooth border (color) lines limiting well defined areas on this plot.

In Fig. 5(a), the RP of the mere log-periodic oscillations (8) is plotted. As can be seen, as the period decreases, the white lines, indicating periodicity, as in Fig. 2(b), are not parallel to the main diagonal anymore but converge toward the upper right corner; this is because the log-periodic oscillations are more frequently occurring near the critical time.

The two contributions (7) and (8) together generate the RP plotted in Fig. 6. The “arrow” shape is due to the trend, as in Fig. 4; the not too smooth border “color” lines are due to the log-periodicity, as seen in Fig. 5.

At this time it could be interesting also to consider a phase transition signature. In Fig. 7(b), an arbitrary signal is plotted before and after a peak, whence near a crash, taking into account some overall shape like that of an anti-bubble after a crash.40 The RP aspect of Fig. 7(a) reveals a feature far from the normal signal evolution, that is the crash time interval. Notice the well marked black bands corresponding to the crash time.

4. Recurrence Quantification Analysis

In 1992, Zbilut and Webber16 proposed to quantify the presence of patterns, like parallel lines of RPs, through statistical values and gave it the name “Recurrence Quantification Analysis” (RQA). As RP, this methodology is independent of limiting constraints, like the data set size, and does not require stationarity, linearity and any usual assumptions on the probability distribution of data. For these reasons, RQA, as the RP, seems very useful for characterizing state changes.

Initially, Zbilut and Webber introduced five variables to quantify changes; for detailed formulas see the Appendix B. The five basic RQA variables are briefly defined as follows with a brief comment on each one:

- **%REC:** the percentage of recurrent points; recall that a point \((i, j)\) is recurrent if the distance between the vectors \(y(i)\) and \(y(j)\) is less than the threshold; in other words %REC is the ratio of the number of recurrent states measured with respect to all possible states.

- **%DET:** the percentage of recurrent points forming line segments parallel to the main diagonal. The presence of these lines reveals the existence of a deterministic structure.
Fig. 4. (a) RP of a signal in (b) as generated by Eq. (7). The arrow shape in plot (a) is the sign of a strong trend.
Fig. 5. (a) RP of a log-periodic signal as generated by Eq. (8), as shown in (b).
Fig. 6. (a) RP of a log-periodic signal as generated by Eq. (1), as shown in (b). The arrow shape in plot (a) is the sign of a strong trend; the curve lines are due to the log-periodicity.
Fig. 7. (a) The RP of a simulated phase transition of a signal (b) following the law (1) before and after the critical event.
MAXLINE: the longest line segment measured parallel to the main diagonal. In Trulla et al., it is claimed that this quantity is proportional to the inverse of the largest positive Lyapunov exponent. A periodic signal produces long line segments, while short lines indicate chaos.

TREND: the slope of line-of-best-fit through $\%REC$ as a function of the displacement from the main diagonal (excluding the last 10% range). This variable quantifies the drift and the nonstationarity of the time series. For example, a flat slope indicates stationarity because in an homogeneous plot the quantity of recurrent points at the left and at the right of the central line is almost the same; instead when on the RHS this quantity is less than the quantity on the LHS. The variable TREND assumes negative values; this means that the system is going far away from the state it had, i.e., it has a trend.

ENT: the Shannon entropy of the distribution of the length of line segments parallel to the main diagonal. The entropy gives a measure of how much information one needs in order to recover the system. A low entropy value indicates that few informations are needed to identify the system, in contrast, a high entropy indicates that much informations are required. The entropy is small when the length of the longest segment parallel to the diagonal is short and does not vary much. This has to be associated with information on determinism. A high entropy is typical of periodic behavior while low entropy indicates chaotic behavior.

All these variables provide information about different aspects of the plot and are in fact inter-correlated.

For RQA computation VRA has been used. It permits to do the computation on selected time intervals of variable sizes. Two complementary studies have been made: in the first case, the RQA variables have been computed for the whole time series, in the other case, we divided the whole time series into subseries and computed the variables in each subinterval, called an epoch. Each epoch is 100 days long and regularly shifted by 10 days, in such a way that each epoch overlaps the next one by 90 days. If $N_e$ is the length of each epoch and $d_e$ the shift, the epoch $i$ corresponds in the time series to the days starting in $t = (i - 1)d_e + 1$ and ending in $t = (i - 1)d_e + N_e + 1$.

The first investigation has been devised in view of comparing global effects due to structures in subseries, while the computation for various epochs has been made to emphasize the changes in state inside the whole time series.

A notice on the displayed data is here in order. For conciseness, the results presented in table forms correspond to the variables computed on the whole time series, while the figures correspond to the variables computed on epochs.
Table 2. RQA of Brownian motion, sine, white noise.

<table>
<thead>
<tr>
<th>Series</th>
<th>Brownian motion</th>
<th>Sine</th>
<th>White noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>%REC</td>
<td>18</td>
<td>15</td>
<td>0.6</td>
</tr>
<tr>
<td>%DET</td>
<td>62</td>
<td>69</td>
<td>~ 0</td>
</tr>
<tr>
<td>MAXLINE</td>
<td>397</td>
<td>997</td>
<td>0</td>
</tr>
<tr>
<td>ENT</td>
<td>4.3</td>
<td>3</td>
<td>~ 0</td>
</tr>
<tr>
<td>TREND (units/1000 points)</td>
<td>−16</td>
<td>−0.8</td>
<td>~ 0</td>
</tr>
</tbody>
</table>

Table 3. RQA of simulated signals: Trend, log-periodicity as in function (8), function (1).

<table>
<thead>
<tr>
<th>Subseries</th>
<th>Trend</th>
<th>Log-periodicity</th>
<th>Function (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>1000</td>
<td>1000</td>
<td>400</td>
</tr>
<tr>
<td>%REC</td>
<td>37.9</td>
<td>19.2</td>
<td>39</td>
</tr>
<tr>
<td>%DET</td>
<td>~100</td>
<td>89.8</td>
<td>99</td>
</tr>
<tr>
<td>MAXLINE</td>
<td>997</td>
<td>987</td>
<td>383</td>
</tr>
<tr>
<td>ENT</td>
<td>8.5</td>
<td>6.0</td>
<td>7.3</td>
</tr>
<tr>
<td>TREND (units/1000 points)</td>
<td>−113.8</td>
<td>−3.9</td>
<td>−245</td>
</tr>
</tbody>
</table>

In Table 2, the values of RQA variables are given for the Brownian motion, the sinusoid and the white noise in order to illustrate the simple cases studied in the RP section, in a quantitative way, while the RQA variables are given for the simulated trend, the log-periodicity and function (1) in Table 3.

As can be seen from the tables, high values of %REC, %DET, MAXLINE and ENT are associated to deterministic processes, while high values of TREND are associated with a not stationary process having a strong trend.

In Fig. 8, RQA variables values computed on epochs are plotted for the simulated phase transition plotted in Fig. 7.

All the variables, studied for epoch, take high values because the signal is (more) deterministic-like. However, when in coincidence with the phase transition, see epoch 49 for %REC, ENT and TREND, and see epoch 51 for %DET, the RQA variables take their lowest values. All these epochs include the critical time. Instead, the MAXLINE assumes the lowest values in epochs 42 and 57. Notice that the first epoch (42) stops before the phase transition while the second (57) starts at the critical time.

5. Data and Methodology

The financial index data used below are the daily closing price of the German index DAX and the American technology index NASDAQ. The period spanned by the
Fig. 8. The values of the RQA variables for the simulated log-periodic signal (Fig. 7) before and after the phase transition. Note that %REC, %DET, ENT and TREND take the lowest values in correspondence of phase transition period, see epochs 49 and 51, while MAXLINE takes the lowest values in epochs 42 and 57, respectively before and at the beginning of the transition period.
Table 4. RQA of Dax on the three studied periods: Bullish corresponding to 5 January 1998 and 8 March 2000; bearish: 8 March 2000 and 23 January 2003; stationary: 28 July 1998 and 29 November 1999. Note that the lowest values are in the period assumed to be stationary.

<table>
<thead>
<tr>
<th>Subseries</th>
<th>Bullish period</th>
<th>Bearish period</th>
<th>Stationary period</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>552</td>
<td>728</td>
<td>341</td>
</tr>
<tr>
<td>%REC</td>
<td>48</td>
<td>54</td>
<td>15.3</td>
</tr>
<tr>
<td>%DET</td>
<td>12.8</td>
<td>9.3</td>
<td>2.8</td>
</tr>
<tr>
<td>MAXLINE</td>
<td>511</td>
<td>687</td>
<td>141</td>
</tr>
<tr>
<td>ENT</td>
<td>5.5</td>
<td>5.7</td>
<td>1.5</td>
</tr>
<tr>
<td>TREND (units/1000 points)</td>
<td>−38.8</td>
<td>−47.1</td>
<td>−31.3</td>
</tr>
</tbody>
</table>

Table 5. RQA of NASDAQ on five the subseries studied of 200 days. The subseries (I) starts in January 1998 and ends in October 1998; the subseries (II) starts in June 1998 and ends in February 1999; the subseries (III) starts in October 1998 and ends in July 1999; the subseries (IV) starts in February 1999 and ends in December 1999; the subseries (V) starts in July 1999 and ends in May 2000. Note that the subseries (IV) takes the highest values.

<table>
<thead>
<tr>
<th>Subseries</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>%REC</td>
<td>6.075</td>
<td>9.141</td>
<td>5.513</td>
<td>17.146</td>
<td>9.246</td>
</tr>
<tr>
<td>%DET</td>
<td>35.980</td>
<td>36.119</td>
<td>29.079</td>
<td>45.018</td>
<td>54.511</td>
</tr>
<tr>
<td>MAXLINE</td>
<td>125</td>
<td>158</td>
<td>83</td>
<td>179</td>
<td>166</td>
</tr>
<tr>
<td>ENT</td>
<td>2.522</td>
<td>3.547</td>
<td>2.585</td>
<td>4.054</td>
<td>3.301</td>
</tr>
<tr>
<td>TREND (units/1000 points)</td>
<td>−105.125</td>
<td>−155.824</td>
<td>−97.808</td>
<td>−273.775</td>
<td>−138.501</td>
</tr>
</tbody>
</table>

The data is January 1998 till November 2003, giving a total of 1493 daily observations for DAX and 1482 daily observations for NASDAQ. They are plotted in Figs. 9(a) and 10(b) respectively. Both time series have been first transformed into the classical momentum divided by the maximum $X_{\text{max}}$ of the time series over the time window which is considered in the specific analysis:

$$X^\text{norm}_i = \frac{X_i - X_1}{X_{\text{max}}}. \quad (9)$$

RPs have been plotted for each time series, RQA variables have been computed on different epochs, see Figs. 11 and 12. The data has been afterwards divided into subseries and several of these analyzed again; see Tables 4 and 5 for the corresponding definitions of such subsets and the RQA variable values.

6. Practical Considerations

At this stage, before embarking into a discussion of the financial index results, practical considerations are to be made. It should be recalled that a search for $d$ and $m$ values must be made first. This is done, in TISEAN, through the function “mutual”
Fig. 9. (a) Daily closing price for DAX from 5 January 1998 to 23 November 2003; (b) RP of DAX from 5 January 1998 to 23 November 2003. The grey band with an “L” shape almost centered around the point of coordinates (530, 200) in plot (b) corresponds to the peak reached in March 2000. The dark grey band on the right centered in \( x = 1300 \) represents the difference between the bullish evolution at the beginning of 2000 and the bearish one in 2003.
Fig. 10. (a) Daily closing price of NASDAQ from 5 January 1998 to 21 November 2003; (b) RP of NASDAQ from 5 January 1998 to 21 November 2003. The dark grey band delimited by horizontal coordinates $x = 504$ and $x = 566$, encircled by a grey area delimited by horizontal coordinates $x = 452$ and $x = 690$, is the “image” of the crash of April 2000. It is a “strong event” but afterwards the normal regime is restored.
Fig. 11. The values of the RQA variables for the daily closing price of DAX in 100 days epochs shifted by 10 days. It is worth to note that, the highest absolute values are in epochs 38/44 and 82/87, corresponding respectively to the period of the highest DAX peak (July 1999 and March 2000) and to the deepest “valleys” (March 2001 and September 2001) reached by DAX in the time span.
Fig. 12. The values of the RQA variables for the daily closing price of NASDAQ in 100 days epochs shifted by 10 days. The highest absolute values are reached in epochs between 37 and 43, also corresponding to the period from 10 June 1999 to 28 January 2000, when the bubble is growing.
Fig. 13. The two figures (up) are of DAX, while the figures (down) are of NASDAQ. (LHS) (up and down) output by TISEAN of the function “mutual”: on the \( x \)-axis, there is the time delay, while on the \( y \)-axis the corresponding values of the mutual information. (RHS) output by TISEAN of the function “false nearest”, on the \( x \)-axis there is the embedding dimension while on the \( y \)-axis the relative percentage of fnn.
which computes the “mutual information” while the function “false nearest” computes the percentage of false nearest neighbor. In Fig. 13, the values of the mutual information and the percentage of FNN for the DAX (top) and for the NASDAQ (bottom) are shown. On the left, the mutual information function is displayed when computed for each time lag \( d \), while on the figure RHS the percentage of false nearest neighbors is shown for each dimension \( m \). Recall that the optimum time delay \( d \) is chosen as the one which minimizes the average mutual information of Eq. (3), while the embedding dimension \( m \) is chosen as the value for which the percentage of false nearest neighbors drops to zero.

Even if it is difficult to see a marked minimum in both Fig. 13 (LHS) (top and bottom), for both DAX and NASDAQ, the value chosen for the time delay is hereby taken equal to 10; while the embedding dimension can be chosen to equal to 5. These values are those used below to reconstruct the space even when the subseries are analyzed.

Consider that each delayed vector is constructed along a period of \((m-1)d\) days, given a set of \( d \) and \( m \) values, with this choice the state space is reconstructed over ca. Two months. In doing so, each vector \( y(i) \) of the reconstructed space somewhat represents the evolution of the market during these two months, with a two weeks lag; this kind of phase space reconstruction can be expected to allow one to recognize changes at an appropriate time. Larger parameters values imply long time intervals and in fact induce less sensitivity to changes on a short time scale.

In the following, RP is further used to have a glance about the behavior of DAX and NASDAQ along the studied time span. The discussion will be such that after observing “epochs” in which the signal could be assumed to be stationary and about finding rare events, both financial index time series are divided into subseries, themselves reconsidered with RP and RQA techniques for finer discussion.

7. DAX

The RP of DAX between 5 January 1998 and 23 November 2003 is shown in Fig. 9(b). It can be first observed that this RP is not homogeneous. In fact, stationarity in such financial time series can hardly be assumed to occur over a period of six years. Let us zoom inside the RP and observe how each band (black, white or grey) can be interpreted. Between coordinates 0 and 430 the RP looks quite white. For this reason, it can be deduced that during the first two years of data no (rare) remarkable event occurred, except for the DAX maximum reached in July 1998, corresponding to the small grey band in the lower left hand corner in Fig. 9(b). The dark grey area with an “L” shape encircled in a clear grey area that is centered around the point of coordinates (530, 200) corresponds to the end of 1999 and the beginning of 2000, when the behavior of DAX changes and the index has an upward acceleration. This feature corresponds to the acceleration ending in March/April 2000. Thereafter a bearish period started, going on until the beginning of 2003. The dark grey area centered around coordinates (1300, 520) represents the
difference between the bearish period at the beginning of 2003 and the bullish one at March 2000.

The values of RQA variables computed on different epochs are shown in Fig. 11 for the DAX. It can be seen that the highest absolute values are found in the epochs 38 to 43, corresponding to July 1999 and March 2000, where an acceleration took place. Other high value epochs are between 82 and 87 corresponding to March 2001 and September 2001, respectively, when the DAX has the deepest descend during the bearish period. Notice that all the RQA variables assume the lowest values in epochs 13 to 35 corresponding to the end of July 1998 and the end of November 1999 when the behavior looks like stationary.

Therefore in view of these observations the DAX time series was divided first into two subseries: one going from 5 January 1998 to 8 March 2000, i.e., the bullish period, the other from 9 March 2000 to 23 January 2003 which represents the bearish period. Moreover, it has to be observed that the white square area on the lower left hand corner in the RP is the “image” of the 28 July 1998 to 29 November 1999 time corresponding in RQA exactly to the epochs having the lowest values. No evident trend is present in this period, whence the corresponding region of RP looks homogeneous. Therefore, the signal in that period can be assumed to be rather stationary and studied separately.

Figure 14 is the RP of the bullish period from 5 January 1998 to 8 March 2000, 552 days; the large white area in the upper right corner is a stationary period, instead on the RP borders the grey areas becoming darker are due to the drift.

![Fig. 14. RP of DAX from 5 January 1998 to 8 March 2000, corresponding to the bullish period. The large white area in the upper right corner is a stationarity period; instead, on the borders, the grey bands becoming darker indicate a drift.](image-url)
Fig. 15. RP of the daily closing price of DAX from 8 March 2000 to 23 January 2003. It is the RP of the bearish period; in the first 350 days, the price has a soft negative trend that corresponds to a white area and a fading grey. The grey area delimited between coordinates \( x = 350/400 \) corresponds to a deep minimum in DAX on 21 September 2001. After this minimum, the price rises a little bit but the negative trend leads to an “arrow” shape, i.e., the white area on the upper right corner centered in (450, 500).

Figure 15 is the RP of the bearish period from 8 March 2000 to 23 January 2003, 728 days; in the first 350 days the price has a soft negative trend that corresponds to the white area and a fading grey. The grey area delimited between coordinates \( x = 350/400 \) corresponds to the deep minimum in DAX on 21 September 2001. After this minimum, the index rises a little bit but then, again a negative trend follows to form an “arrow” shape; the dark grey area on the lower right corner shows the index difference in value just after the crash and after almost two years.

Figure 16 is the RP for a stationary signal period from 28 July 1998 to 29 November 1999, 341 days. The RP looks quite homogeneous, there is a small black band on the lower left corner, but the stationarity is quite proven. The RQA variable values for these time series are given in Table 4. It is worth to note, that the period considered as stationary is that with the lowest values, as already seen in RQA results when the technique is applied to distinct epochs.

8. NASDAQ

Figure 10(b) is the RP of NASDAQ between 5 January 1998 and 21 November 2003. Of interest is the dark grey vertical band surrounded by a lighter grey area. The grey area is delimited by horizontal coordinates \( x = 452 \) and \( x = 690 \), the dark grey one is delimited by horizontal coordinates \( x = 504 \) and \( x = 566 \).
The grey band corresponds to 19 October 1999 to 27 September 2000; the dark grey band corresponds to 3 January 2000 to 31 March 2000. The dark grey area suggests that those days are “far away” from all other days in a phase space sense. In fact, all white regions represent recurrence, while the black or grey areas represent rare events (see discussions in Sec. 3). Note the same RP shape of the phase transition simulation in Fig. 7 and note also that the dates here above fall in the same time interval as the bubble and the subsequent crash.

In order to have more insight, let us observe that in contrast to the RP of DAX only one event emerges strongly from the plot, i.e., the crash.

Thus, it can be supposed that the initial bubble time occurs at \( x = 452 \) (19 October 1999), i.e., the coordinate of the border line of the grey area in Fig. 10(b); let us call it \( t^* \). We can thus deduce that on such a day the evolution of the system changes, i.e., the evolution passes from a normal regime to a critical regime, as at the Ginzburg temperature\(^{42}\) in phase transition studies.

In order to further analyze whether and how the data changes, the series has been divided into subseries of 200 days shifted by 100 days for each analysis, such that each subseries is overlapping during almost five months.

For the first subseries (I) starting in January 1998, and ending in October 1998, the RP in Fig. 17 presents a small dark grey area at \( x = 130 \), when the NASDAQ...
Fig. 17. RP of the NASDAQ subseries (I) starting in January 1998 and ending in October 1998; the dark grey small area near $x = 130$ is the maximum NASDAQ price around 2000, in July 1998; except this the remaining part of the signal can be assumed stationary because the RP looks homogeneous.

Fig. 18. RP of the NASDAQ subseries (II) from June 1998 to February 1999; this RP is quite homogeneous except for some isolated not recurrent points aligned along the main diagonal.
reached prices around 2000 in July 1998. The remaining part of the RP looks homogeneous. Notice that RQA variables in Table 5 do not show any remarkable value.

The second subseries (II) goes from June 1998 to February 1999: the RP in Fig. 18 is quite homogeneous except for some isolated not recurrent points aligned along the main diagonal. The RQA variables in Table 5 are a little bit larger than for I.

The third subseries (III) is from October 1998 to July 1999; the RP in Fig. 19 looks quite homogeneous except for some regions corresponding to local peaks. The RQA variables are smaller than for I and II.

The fourth subseries (IV) goes from February 1999 to December 1999; the RP in Fig. 20 shows the characteristic shape, typical of a strong trend as studied and pointed out in Fig. 4(a) and 6(a). The trend starts to be significant in the middle of October 1999. This indicates that the RP has changed indeed when the bubble has started. The RQA variables in Table 5, evidence in this period, the highest values. It has to be underlined that this IV period does not include the crash time, but stops before the bubble bursts.

It is worth to emphasize the similarity between Figs. 6 and 20: it indicates a pattern in IV similar to the simulation of the log-periodic function (1).

The fifth V subsseries goes from July 1999 till the end of May 2000. The RP shown in Fig. 21 indicates a grey area which starts near \( t^* \), in the plot corresponding with \( x = 48 \). However, the bubble is not so evidently seen in this RP, as it was for the time interval studied above.
Fig. 20. RP of the NASDAQ subseries (IV) from January 1999 to December 1999. This RP assumes the characteristic "arrow" shape typical of a strong trend: it starts to be significant in the middle of October 1999. It has to be underlined that this period does not include the crash time, but stops before the bubble bursts. It is worth to note also a similarity between this RP and the RP in Fig. 6, i.e., the simulation of the log-periodic function (1).

Fig. 21. RP of the NASDAQ subseries (V) from July 1999 to May 2000; the grey area starts when the bubble starts; however, in this RP, the bubble is not so evident because it is not possible to see the change between normal and critical regimes as in Fig. 20.
Fig. 22. (a) The NASDAQ before the crash detrended by function (7) (the smooth line) (b) its RP. It was expected to see signs of the log-periodicity but, due to the presence of noise, we are not allowed to affirm so.
It has been also tried to detrend the data by the function (7), see Fig. 22(a). The difference of the data and the fitted function produces an oscillating series whose RP is shown in Fig. 22(b).

9. Conclusion

One of the main aims of this review was, first of all to indicate how to use RP and RQA in order to distinguish normal regimes from critical regimes. The case of financial index data series was used as an illustration. The techniques were applied to DAX and NASDAQ between January 1998 and November 2003, encompassing the crash of April 2000, i.e., presenting a bullish period followed by a bearish one.

As seen, the RP and RQA results evidence very well defined differences in the state and evolution of signals when comparing data in different epochs. Indeed, from an RP one can argue about the stationarity and the behavior of the series when looking at the homogeneity and at the different “color bands”. From the values of RQA variables it can be argued about the deterministic behavior of the signal. Indeed, if the RP does not look homogeneous and if black or dark grey bands are present, it can be argued that a rare event occurred, as seen in the simulation in Fig. 7(a). Moreover, at the same time the RQA variables, in particular %DET and ENT, take high values. In general, the low values of these variables can be associated to stochasticity, while high values are associated with a deterministic behavior, as seen in Table 2. The inhomogeneity of RP and the high values of RQA variables reveal events that can be interpreted as a thermodynamic phase transition, if the homogeneity and low values are restored after this event.

Turning toward specific data analysis, DAX and NASDAQ, looking at RPs in Figs. 9 and 10 one fact is evident: the same “rare” event has occurred at the same time (around coordinates 500) in both markets. However, this event is better emphasized through the grey band in the RP of the NASDAQ than in the RP of the DAX. In fact, in Fig. 10 the grey band representing this event is quite isolated, suggesting that after this event a normal regime was restored, i.e., like when a phase transition occurs, see again Fig. 7. In Fig. 9, the same event is not so obvious but it can be still seen as a strong fluctuation, because in the DAX the event is more like an extended drop than a real crash.

Each coordinate in RP is linked with the time series, so the border line of a grey or black band reveals the time when the data behavior starts to change like for a thermodynamic transition. As seen in the previous section, the border line of the grey band around the black one was delimited by coordinates 452 and 690 that correspond to the 19 October 1999 and 27 September 2000 days. This is an a posteriori estimation, but through the analysis of the fourth subseries IV, one can argue to have found the beginning of the bubble, and have an estimate, with some delay, on the initial moment before the bubble grows, that in our cases corresponds to 16 December 1999. This delay consists in \((m - 1)d\) days (ca. 120 days). The fourth subseries IV goes from 16 March 1999 to 28 December 1999. It includes the
initial moment and gives the range of time necessary for the estimation. Notice that the subseries does not contain the crash time, but yet the features warn about its occurrence.

It is worth also to note the similarity between Figs. 6 and 20. Starting from the point of view of log-periodic oscillation, the log-periodic model in Eq. (1) was simulated and the generated signal was studied by RP. The resulting plot was compared with the RP of the fourth NASDAQ time series. Some similarity can be seen in them, the arrow shape is due to a strong trend, see Fig. 4; it has been seen that a rather smooth trend leads to a “rather smooth” plot, while the effect of the log periodicity leads to border color line curves with projections like in Fig. 6. Nobody can expect from such an experimental data to have a truly smooth trend, in fact, Fig. 20 has no smooth border line. Some of those curve lines could be due to the log periodicity but the presence of noise does not allow us to affirm it. The same consideration holds also for Figs. 7(a) and 10(b), and also for Figs. 5(a) and 22(b).

In conclusion, the possibility to use the RP and RQA methods in order to detect a bubble regime and to find the initial bubble time (by analogy to the Ginzburg temperature\(^{42}\)) has been investigated. It has been shown that, with some delay \((m - 1)d\) days) as respect to the beginning but enough time before the crash (three months in this particular case), such that a warning could be given, RP and RQA detect a difference in state and recognize the critical regime.

Acknowledgments

The authors thank COSTP10 for an STSM grant to A. Fabretti and acknowledge discussions with Claire Gilmore, Charles Webber and Giulia Rotundo.

Appendix A. Interpreting RP

The following scheme is a summary of what has been discussed in Sec. 3.

<table>
<thead>
<tr>
<th>Homogeneity</th>
<th>Stationarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Darkening from the upper left corner to the lower right corner</td>
<td>Nonstationarity: drift and trend</td>
</tr>
<tr>
<td>Black or dark grey bands</td>
<td>Nonstationarity: some system states are far from the medium evolution or can correspond to a transition.</td>
</tr>
<tr>
<td>Periodic patterns</td>
<td>There are cycling evolutions of the system; the distance between periodic patterns is the period.</td>
</tr>
<tr>
<td>Isolated recurrent points</td>
<td>Rapid fluctuations of the system,</td>
</tr>
</tbody>
</table>
Diagonal lines parallel to the main diagonal

which could be random; in fact, white noise looks like a set of isolated random points.

Vertical and horizontal lines

The system evolution is similar at different times; the system could be deterministic.

\begin{align*}
\text{Diagonal lines parallel to the main diagonal} & \\
\text{Vertical and horizontal lines} & \\
\end{align*}

\begin{align*}
\text{Diagonal lines parallel to the main diagonal} & \\
\text{Vertical and horizontal lines} & \\
\end{align*}

\begin{align*}
\text{Diagonal lines parallel to the main diagonal} & \\
\text{Vertical and horizontal lines} & \\
\end{align*}

Appendix B. Recurrence Quantification Variables

Let us define

\[ R_{ij} = \begin{cases} 1 & \text{if } (i, j) \text{ is recurrent} , \\ 0 & \text{otherwise} , \end{cases} \tag{B.1} \]

and consider that

\[ D_{ij} = \begin{cases} 1 & \text{if } (i, j) \text{ and } (i+1, j+1) \text{ or } (i-1, j-1) \text{ are recurrent} , \\ 0 & \text{otherwise} . \end{cases} \tag{B.2} \]

In terms of \( R_{ij} \) and \( D_{ij} \), the RQA variables are easily written as follows.\(^4\)

The percentage of recurrent points in an RP is

\[ \% \text{REC} = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{ij} . \tag{B.3} \]

The percentage of recurrence points which form diagonal lines is

\[ \% \text{DET} = \frac{\sum_{i,j=1}^{N} D_{ij}}{\sum_{i,j=1}^{N} R_{ij}} . \tag{B.4} \]

Let \( N_l \) be the number of diagonal lines and let be \( l_i \) the length of the \( i \)th diagonal line. The length of the longest diagonal line is

\[ \text{MAXLINE} = \max(l_i; i = 1, \ldots, N_l) . \tag{B.5} \]

The Shannon entropy for a function \( f(x) \) is

\[ H(x) = - \sum_{x \in \text{dom}(X)} f(x) \log f(x) , \tag{B.6} \]

while for a distribution it is

\[ H(p_1, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log(p_i) . \tag{B.7} \]

Let us denote \( p(l) \) the distribution of the diagonal line lengths, its Shannon entropy is

\[ \text{ENT} = - \sum_{i=1}^{N_l} p(l) \ln p(l) . \tag{B.8} \]
Let us define
\[ RR_i = \sum_{j=1}^{\tilde{N}} R_{ij}, \]  
(B.9)
where \( \tilde{N} \) is the number of recurrent points after rejecting 10% of the farthest away occurring points; in other words \( RR_i \) is the number of recurrence points with respect to the state \( i \). The formula for the trend obtained minimizing the least square error is
\[ \text{TREND} = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(RR_i - \overline{RR}_i)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2}, \]  
(B.10)
where \( \overline{RR}_i \) denotes the average of \( RR_i \) over \( i \in [1, \tilde{N}] \).

References