Visual recurrence analysis as an alternative framework for time series characterisation

Branko Pecar
University of Gloucestershire, UK
bpecar@glos.ac.uk

Abstract

Basic recurrence plot analysis has been extensively used as a technique for characterising financial time series. In this paper we examine recurrence plots of two of the most fundamental processes, i.e. the white noise process and the Wiener process. These recurrence plots are also used to 'catalogue' typical behaviour for different time lags for these two processes. The graphs are then used as a template to compare the recurrence graphs of the returns on eight real-life data sets taken from the NYSE.
Keywords: VRA, visual recurrence analysis, recurrence plots, non-linear time series analysis, deterministic systems

1 Introduction

Rescaled range analysis and recurrence plots are some of the techniques that have been introduced over recent years for characterising financial time series. Today, primarily thanks to various software packages freely available from the Internet, these techniques are more accessible than ever before. In this paper we used a package called VRA (Visual Recurrence Analysis), created by Eugene Konov [5]. Interpretation of the results is somewhat arbitrary, making it more of a qualitative than a quantitative tool. However, the rationale behind such an analysis is highly rigorous and offers a number of benefits obscured by other more 'quantitative' techniques. The intentions of this paper are to explore whether, by extending the knowledge of time series characteristics, the results gained could later on be exploited to produce more accurate forecasts.

Recurrence analysis, as a method for identifying hidden patterns and revealing non-stationarity, was introduced by Eckmann, Kampshort and Ruelle [2]. The method is used for visualising and studying system phase space orbits. The logic of the method is very intuitive. Dynamical systems are often represented as series of differential equations, each containing variables that can define the state of such a system. However, in real life the best we can often do is to get a single function describing the movements of a system. Despite the fact that state variables are not known, by calculating the derivatives of this single function, we can create equations of motion. These equations generate the space that is related to the original space where the actual system resides. Takens [11] proved that the whole process could be simplified even further. The dynamics on the attractor of the underlying system are topologically equivalent to those of a state space system created from a single observable variable. This means that it is possible to reconstruct a topologically equivalent picture of the entire dynamical system (hidden and otherwise unobservable multidimensional space) from a single time series. The way of achieving this is via embedded delayed coordinates.

2 Objectives

This paper intends to provide a brief overview of how recurrence plots are rendered and how colours can be used to construct a richer visual environment for interpreting time series. Several well-described processes will be catalogued on an ad hoc basis, depending on the characteristics of the embedding space and the corresponding visual appearance. This approach to analysis will be extended to several real-life financial time series, namely closing minute and daily stock values from New York Stock Exchange (NYSE). The intentions are to establish whether coloured recurrence plots can be used to classify and characterise time series.

In order to facilitate this characterisation, both the minute and daily, as well as short and long stock returns, will be compared. Their recurrence plots will be examined and compared with the ad hoc classification of the systems previously analysed. The ultimate objective is to establish whether visual recurrence analysis can provide a clue regarding the presence of deterministic, possibly even chaotic, movements in financial time series.

3 Recurrence plots

In order to create a phase space from a single variable, observations from the time series need to be turned into vectors Y_t. This is easily achieved in the following fashion:

$$Y_t = (x_t, x_{t+\tau}, x_{t+2\tau}, \ldots, x_{t+(m-1)\tau}) \quad (1)$$

Where,

x_t = time series observations
t = units of time
τ = time delay (sampling time)
m = embedding dimension

A sufficient number of vectors Y_t create a phase space that is topologically equivalent to the original, unobservable multidimensional space that hosts the underlying system. In a way, the observed time series is just a projection (an output) of this underlying process (system) into a time window. Once the system has been reconstructed it is possible to analyse the properties of such systems, otherwise hidden in the time window.

The real difficulty is in deciding how to define the vectors that create this space. From the formula in eqn. (1), we can see that only two parameters will define every vector. These are *m*, the embedding dimension and τ, the time delay. Both parameters need to be estimated as precisely as possible. A reconstructed space, in order to be representative, must have the correct embedding dimension, more specifically this is only possible if $m \geq 2n+1$, where n represents a number of variables (or degrees of freedom) of the dynamical system [1]. The correct value of *m* is established via the false nearest neighbour method (for details see [9]). Incorrect selection of *m* leads towards a space where the topological structure is no longer preserved. The time delay τ (sampling time) serves the same purpose as the embedding dimension and an incorrectly defined sampling time will affect the reconstruction of the attractor. The simplest way to determine the optimal value of τ is either via the autocorrelation function or, preferably, the mutual information function [4].

Once the vectors are calculated, a space can be reconstructed and a number of system invariants can be measured (the correlation dimension, maximum Lyapunov exponent, etc.). However, rather than taking this analytical approach, a simple recurrence plot can be constructed. Recurrence plots display distances between vectors and, if those distances are below some threshold value, the point is marked. Depending on the application, the distances can be shaded, or colour coded. Such plots provide a very intuitive and easy to use tool for studying the motion of the system trajectories. A general rule that applies is, the more deterministic the signal, the more structured the recurrence plot will be.

4 Time series description

To explore the descriptive and analytic capabilities of this approach to analysis, eight real-life financial time series and four artificially generated series were selected. The real life series are stock movements taken from the New York Stock Exchange. The series are daily and minute closing values for BP, IBM, Coca Cola and Pfizer (Tickers symbols: BP, IBM, KO and PFE). The details for all eight series are given in Table 1. All minute data cover Monday to Friday intra-day trading from 09:30-16:00.

Table 1: Four financial time series from the NYSE selected for analysis

Stock values		BP	IBM	Coca Cola	Pfizer
Daily closing values					
	Start	3-Jan-1977	2-Jan-1962	2-Jan-1970	4-Jan-1982
	Finish	9-May-2003	9-May-2003	9-May-2003	9-May-2003
No of observations		6653	10405	8423	5391
Minute closing values					
	Start	22-Apr-2003	22-Apr-2003	22-Apr-2003	22-Apr-2003
	Finish	5-May-2003	5-May-2003	5-May-2003	5-May-2003
No of observations		5458	5862	5473	5473

However, rather than analysing the closing values of the stocks, the return values were calculated. If a stock price, at any point in time, is defined as X_t and Δ_t is a change in time, then the price change is calculated as:

$$R_t = \frac{X_{t+\Delta t} - X_t}{X_t} = \frac{Z_t}{X_t} \qquad (2)$$

As $Z_t/X_t \approx \log X_{t+\Delta t} - \log X_t$, if Z_t/X_t is small, the value of R_t can be more efficiently expressed as:

$$R_t = \log X_{t+\Delta t} - \log X_t \qquad (3)$$

This was the method adopted for calculating returns. Specific returns calculated in this paper applied to lags 1 (short returns) and 1,000 (long returns). In the context of daily returns, 1,000 lags covers a period of 2.7 years and in the context of minute returns, 1,000 lags covers 16.7 hours, i.e. approximately 2.5 trading days.

The four artificially generated series were: a white noise process, a Wiener process (Brownian motion with Markov characteristics, often called fractional Brownian motion), a sinusoid and the Lorenz attractor. For all four series more than 5,000 observations were generated. The white noise process was generated on Excel using the RAND() function. The Wiener process was generated as an integrated white noise process. The sinusoid series that was generated was periodic every 101 observations and the Lorenz attractor was generated using difference equations (only 'x' variable evolving in a time window has been used) with the following parameter values:

$$dx/dt= -10(x +y) \qquad (4)$$
$$dy/dt= 28x -y -xz \qquad (5)$$
$$dz/dt= -8/3z +xy. \qquad (6)$$

For these four simulated series, artificial ‘returns’ (i.e. differenced values) were also calculated for identical time lags as for the stock exchange data, i.e. for 1 and 1,000 lags.

5 Recurrence plot analysis

The closeness of vectors in the phase space was shaded in accordance with the scheme shown in Fig. 1. Distances displayed as 0.00 for white and 0.04 for black varied from series to series and from global to local measurements. The metric used was the Euclidean distance.

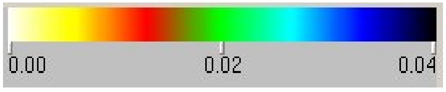


Figure 1: Colour map for marking vector distances on the recurrence plot

According to this map, the closer the two vectors on the recurrence plot, the whiter the shading. The white diagonal line, visible on some plots, represents the line of identity, i.e. every vector measured against itself. The recurrence plot, therefore, consists of two mirroring sections divided by the identity line that flows from the bottom left to the top right corner.

Fig 2a below shows a typical recurrence plot for a white noise process and Fig 2b a typical random walk (Wiener process) recurrence plot. Both plots, although rich in texture in a different way, represent typical visual signatures of these two processes. However, as expected, neither of them appears to contain any regularity.

The recurrence plot of the Lorenz attractor in Fig. 3a, on the other hand, shows some regularity. These regularities are also preserved if very long differences are calculated. The sinusoid in Fig 3b shows exactly the same characteristics. As a matter of fact the plot for the original time series is indistinguishable from the plot for the series containing long differences.

a. b.

Figure 2: A White noise (a) and the Wiener process (b)

It was indicated earlier that plots based on the correct time delayed embedded coordinates might produce patterns that are not normally visible in the series and which could provide clues to search for the presence of deterministic chaos. To illustrate this, Figs 4a and 4b display recurrence plots for the Lorenz attractor and the sinusoid, each presented at the appropriately calculated level of *m* and τ.

a. b.

Figure 3: The Lorenz attractor (a) and a sinusoid (b)

As we can see, the plots enhance the features we anticipated to find. The Lorenz attractor shows a stream of light lines parallel with the identity line. These lines indicate the presence of unstable periodic orbits. The sinusoid plot clearly confirms the periodicity of the sinus waves, which corresponds with every light line parallel with the identity line (the first light line parallel with the identity line starts at vector 101, corresponding with the periodicity of the original sinusoid).

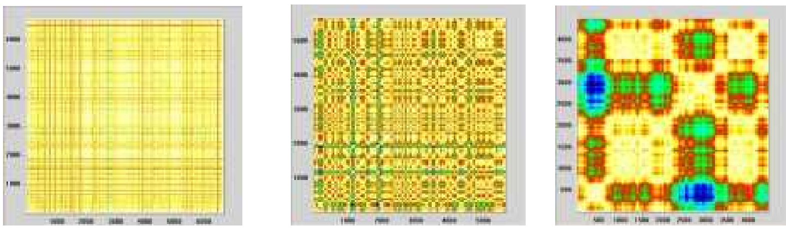
a. b.

Figure 4: Lorenz attractor $m=2, \tau=16$ (a) and Sinusoid $m=2, \tau=25$ (b)

6 VRA plot analysis

To illustrate the difference, recurrence plots for the financial time series that we selected from the NYSE were produced. Fig. 5. No matter which series were picked, the plots looked very similar. The first plot (Fig. 5a) is typical of all eight examined short return (lag 1) series. The middle plot (Fig. 5b) is a typical 100 lag return series and the right one (Fig. 5c) is a characteristic 1,000 lag return series. Interestingly, the plots are very similar for both daily and minute returns, confirming that the financial series returns seem to be insensitive to sampling frequency.

What is also characteristic is that the first plot resembles very much a recurrence plot for the first differences of the white noise process, whilst the last one is very much like a Wiener process. This is very much in line with the fact that longer horizon returns converge to Gaussian, as predicted by the central limit theorem [6]. The middle plot represents a transition from a short-term (white noise) to a long-term (Wiener process) memory process. If we calculated the Hurst exponent for these series, the estimates would completely support these impressions. These findings, using a different tool set, were also reported by Stanley, although in the context of different power laws governing high frequency vs. low frequency data [10].



a. *b.* *c.*

Figure 5: Recurrence plots for the return series: lag 1 (a), lag 100 (b) and lag 1,000 (c)

Despite the obvious resemblance of the real life financial time series and a white noise process on the one hand and a fractional Brownian motion on the other hand, there is always a possibility that the series are not random, but some kind of a deterministic model masquerading as random [12]. To explore this possibility, we first needed to estimate the appropriate time delays and embedding coordinates, and then zoom in on each and every one of the recurrence plots.

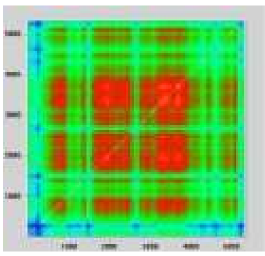
Before the set of ten NYSE financial series were analysed, we produced an ad hoc catalogue for a generic white noise process and a Wiener process presented at different levels of the time delays and embedding dimensions. Both processes were lagged by 1 and by 1,000 observations before the optimum value of the time delay and the embedding dimension for all six time series (three for white noise and three for Wiener process) were estimated. The catalogue with characteristic visual signature of such a space is briefly outlined in Appendix 1. A catalogue for the eight NYSE financial series is given in Appendix 2.

7 VRA plots for financial time series

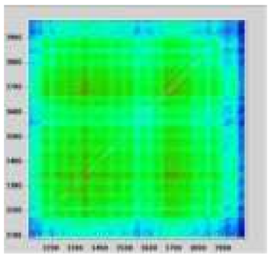
The behaviour of all eight financial time series seemed to follow, remarkably closely, the behaviour of the two most fundamental types of series, i.e. the white noise process and the Wiener process. All the short-term returns, regardless of whether the resolution, i.e. minute or daily returns, behaved as a white noise process. All the long-term returns, again regardless of the resolution, i.e. minute or daily returns, behaved as a proper random walk process (a Wiener process).

In order to discover possible hidden patterns, or the presence of an underlying dynamical system, we needed to look closely at these recurrence plots, looking for short lines parallel with the identity line, signifying unstable periodic orbits. Unfortunately, long term returns do not look promising, i.e. zooming in on a 'Rorschach blob' does not reveal any finer or deeper structure hidden in it. However, the short-term returns offer some tentative hope.

As an example we took daily BP short returns and calculated optimal m=60 and τ=2 (Fig 6a). After zooming in on the section of this recurrence plot (Fig 6b), we can see a number of short lines parallel with the line of identity



a.

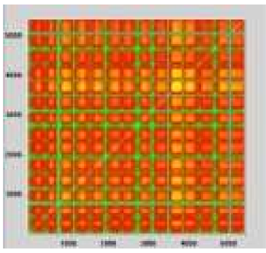


b.

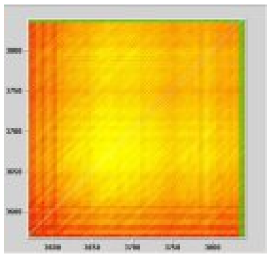
Figure 6: BP BP daily returns τ=2, m=60 (a) and magnification of a section of Fig. 7a (b)

Figs 7a and 7b show similar picture. This is a recurrence plot for PFE single minute returns at m=39 and τ=2. These plots indicate a possibility of unstable periodic orbits embedded in these time series.

We have to point out that in all these cases (and we are arbitrarily showing only two) the lines running in parallel with the identity line are not always as light as one would ideally expect them to be, indicating that recurrence is somewhat tentative and a little more remote than desirable.



a.



b.

Figure 7: PFE minute returns τ=2, m=39 (a) and magnification of a section of Fig 8.a. (b)

The second important point is that these parallel lines do not necessarily describe unstable periodic orbits. It is true that in general the short parallel segments mirror the deterministic character of the system and indicate the presence of unstable periodic orbits embedded in the chaotic attractor of a deterministic system [3]. However, at present the idea that short-term stock exchange returns could possibly be deterministic systems must remain a speculation. A deterministic system might contain non-linear components, but nonlinearity does not have to be reflected in a specific data set [7], and vice versa.

Even better explanation is offered by [8] whose interpretation of financial time series is that they could be best described as a nonlinear noise driven dynamical system far from equilibrium undergoing bifurcation or change in system dynamics. The problem with this approach is that although system dynamics can be well modelled using nonlinear dynamic modelling, this will not help us to produce better predictions. To put it mildly, we are back to square one.

8 Conclusion

Using several artificially generated time series we explained what patterns are likely to be observed in recurrence plots and how they could be catalogued. These findings were used to analyse eight real-life stock exchange time series, representing the minute and daily returns for four data sets from the NYSE. The analysis was extended to include short and long return values.

The long-term returns, for both minute and daily stock values, exhibit behaviour very much in line with that expected from a conventional random walk (Wiener process or fractional Brownian motion). The short-term returns, again for both minute and daily stock values, exhibit behaviour that resembles the white noise differences. However, after estimating appropriate time delays and embedding dimensions, the short-term return series reveal faint regularities, indicating that there is a possibility of some kind of deterministic process driving them. However, without testing this hypothesis by one or more rigorous tests, this remains a speculation. Nevertheless, visual inspection confirms what other more comprehensive studies have alluded to, and that is: a possible presence of nontrivial deterministic nonlinearity in data.

Visual recurrence analysis, in the form of coloured recurrence plots, is a powerful descriptive tool. It is intuitive, quick and a robust method for tentatively classifying and characterising the time series. It can be used as an ad-hoc tool to set up hypotheses that can, later on, be tested with more rigorous tools.

References

[1] Baker G.L., Gollub J.P., *Chaotic Dynamics*, Cambridge University Press, Cambridge, p. 256, 1996.

[2] Eckman J.A., Kampshort S.O., Ruelle D., Recurrence Plots of Dynamical Systems, *Europhysics Letters*, 4 (9), pp. 973-977, 1987.

[3] Holyst J.A., Zebrowska M., Urbanowicz K., Observations of Deterministic Chaos in Financial Time Series by Recurrence Plots, Can One Control Chaotic Economy?, *The European Physical Journal*, B (20), pp. 531-535, 2001.

[4] Kantz H., Schreiber T., *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, p. 304, 2000.

[5] Konov E., VRA Software Package, Version 4.2, accessed 2003, http://home.netcom.com/~eugenek/download.html

[6] LeBaron B., Volatility persistence and apparent scaling laws in finance, accessed 2003, http://212.67.202.199/~msewell/volatility/LeBa99.pdf

[7] Schreiber T., Schmitz A., Surrogate time series. *Physica D* (142), pp. 346-382, 2000.

[8] Small M., Tse C.K., Determinism in financial time series, *Studies in Nonlinear Dynamics and Economics*, 7 (3), Article 5, 2003.

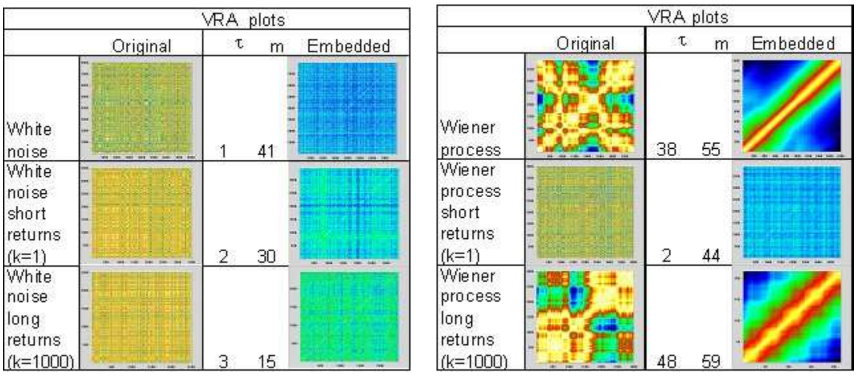
[9] Sprott J.C., *Chaos and Time Series Analysis*, Oxford University Press, Oxford, p. 507, 2003.

[10] Stanley E.H., Exotic statistical physics: Applications to biology, medicine and economics, *Physica* A 285, pp. 1-17, 2000.

[11] Takens F., Detecting strange attractors in fluid turbulence (Chapter). *Dynamical Systems and Turbulence, Warwick, 1980*, ed. Dold, A., Eckmann B., Springer Verlag, Berlin, pp. 366-381, 1981.

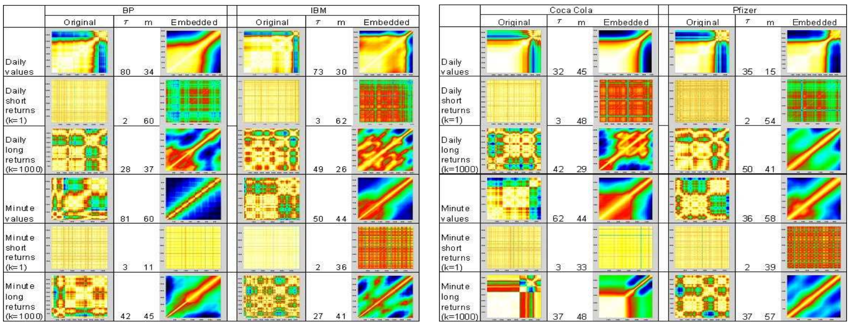
[12] Zbilut J.P., Giuliani A., Webber C.L., Recurrence quantification analysis as an empirical test to distinguish relatively short deterministic versus random number series, *Physics Letters* A 267, pp. 174-178, 2000.

Appendices



a. b.

Appendix 1. VRA plots for three white noise (a) and Wiener processes (b) with optimum time delay level and embedded dimension



a. b.

Appendix 2. VRA plots for various BP and IBM stock values (a) and Coca Cola and Pfizer stock values (b). Plots portray the original time series and the embedded time series with appropriate time delay and embedding dimension.