Nonlinear Dynamics and Recurrence Plots for Detecting Financial Crisis

Peter Martey Addo, Monica Billio, Dominique Guegan

▶ To cite this version:

Nonlinear Dynamics and Recurrence Plots for Detecting Financial Crisis

Peter Martey ADDO, Monica BILLIO, Dominique GUEGAN

Nonlinear Dynamics and Recurrence Plots for Detecting Financial Crisis

Peter Martey ADDO[a,b,c,*], Monica BILLIO[b], Dominique GUÉGAN[c,d]

[a]*European Doctorate in Economics–Erasmus Mundus (EDEEM)*
[b]*Università Ca'Foscari of Venice, Department of Economics*
[c]*Centre d'économie de la Sorbonne (CES) - CNRS : UMR8174 - Université Paris I - Panthéon Sorbonne*
[d]*Ecole d'Économie de Paris - Paris School of Economics (EEP-PSE)*

Abstract

Identification of financial bubbles and crisis is a topic of major concern since it is important to prevent collapses that can severely impact nations and economies. Our analysis deals with the use of the recently proposed 'delay vector variance' (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using wavelet-based surrogates. We exploit the concept of recurrence plots to study the stock market to locate hidden patterns, non-stationarity, and to examine the nature of these plots in events of financial crisis. In particular, the recurrence plots are employed to detect and characterize financial cycles. A comprehensive analysis of the feasibility of this approach is provided. We show that our methodology is useful in the diagnosis and detection of financial bubbles, which have significantly impacted economic upheavals in the past few decades.

Keywords: Nonlinearity analysis, Surrogates, Delay vector variance (DVV) method, Wavelets, financial bubbles, Embedding parameters, Recurrence plots
JEL: C14, C40, E32, G01

1. Introduction

In real-world applications of financial time series analysis, the process underlying the generated signal, which is the time series, are *a priori* unknown. These signals usually contain both linear and nonlinear, as well as deterministic and stochastic components, yet it is a common practice to model such processes using suboptimal, but mathematically tractable models. In general, performing a nonlinearity analysis in a modeling the time series can lead to a significant improvement of the quality of the results, since it facilitates the selection of appropriate processing methods, suggested by the data itself. It is, therefore, essential to investigate the intrinsic dynamical properties of financial time series in terms of its deterministic/stochastic and nonlinear/linear

*Correspondence to: Université Paris1 Panthéon-Sorbonne, MSE-CES UMR8174 , 106-113 boulevard de l'hopital, 75013, Paris, France
Email address: `peter.addo@univ-paris1.fr` (Peter Martey ADDO)

components reveals important information that otherwise remains not clear in using conventional linear methods of time series analysis. For this purpose, we adopt to the recently proposed phase space based 'delay vector variance' (DVV) method (Gautama et al., 2004a), for signal characterisation, which is more suitable for signal processing application because it examines the nonlinear and deterministic signal behaviour at the same time. This method has been used for understanding the dynamics of exchange rates (Addo et al., 2012b), qualitative assessment of machine learning algorithms, analysis of functional magnetic resonance imaging (fMRI) data, as well as analysing nonlinear structures in brain electrical activity and heart rate variability (HRV) (Gautama et al., 2004b). Optimal embedding parameters will be obtained using a differential entropy based method proposed in Gautama et al., 2003, which allows for simultaneous determination of both the embedding dimension and time lag. Surrogate generation used in this study will be based on both the Iterative Amplitude Adjusted Fourier Transform (iAAFT) (Schreiber and Schmitz (1996, 2000)) and a recently refined iAAFT with a wavelet-based approach, denoted WiAAFT (Keylock, 2006).

Recurrence Plot (RP) is an advanced graphical technique of nonlinear data analysis which reveals all the times when the phase space trajectory of the dynamical system visits roughly the same area in the phase space. Although, most applications of this technique has been in physics and biology, it has gained interest in a variety[1] of scientific fields (Marwan et al. (2007)). Recurrence plots and recurrence quantification analysis (Zbilut and Webber (1992); Marwan et al. (2002)) is gaining some attention in financial studys (Kyrtsou and Vorlow (2005); Strozzi et al. (2008); Belaire-Franch (2004); Holyst et al. (2001), among others) .This technique is applicable to any time series since it requires no a prior assumptions on the statistical properties, such stationarity, or mathematical structure of the time series. We find this technique as a promising means of analyzing financial data since it is robust against non-stationarity in the data. In this work, we will mainly use *distance plots*, which is sometimes called unthresholded recurrence plot in studying the dynamics of stock markets.

In this article, our aims is to characterize and detect nonlinear schemes for the US financial cycle considering both S&P 500 Index and Nasdaq Composite. Firstly, our analysis deals with the use of the recently proposed 'delay vector variance' (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using wavelet-based surrogates. This enables the characterisation of the financial time series without a priori assumption on the dynamics or statistical properties of the series. Secondly, we exploit the concept of recurrence plots on both series to locate hidden patterns, non-stationarity, and to examine the nature of these plots in events of financial crisis. In particular, we show the usefulness of recurrence plots in the diagnosis and detection of financial bubbles and crisis, which have significantly impacted economic upheavals in the past few decades.

The paper is organised as follows: surrogate generation methodology and differential entropy based method for determining optimal embedding parameters of the phase-space representation of time series are presented in 2.1 and 2.2 respectively. In section 2.3, an overview of the 'delay vector variance' method with illustrations. We provide in section 2.4, the concept of recurrence plots. In section 3, we present a comprehensive analysis of the feasibility of this approach to analyse the US financial cycle. Section 4 concludes.

---

[1]A comprehensive introduction and bibliography about on recurrence plots is available on http://www.recurrence-plot.tk/ and Marwan et al. (2007)

2

2. 'Delay Vector Variance' Method and Recurrence Plots

In this section, we present an overview[2] of different existing methods successfully applied in physics and engineering. In particular, we show the usefulness of these methods over other methods and then explain how we merged these methods to studying the dynamics of financial time series. Our methodology encompasses surrogate generation methods, differential entropy method for determining the optimal embedding parameters in phase-space, the DVV method and data analysis with recurrence plots.

2.1. Surrogate Data Method and Statistical testing

Surrogate time series, or 'surrogate' for short, is non-parametric randomised linear version of the original data which preserves the linear properties of the original data. For identification of nonlinear/linear behavior in a given time series, the null hypothesis that the original data conform to a linear Gaussian stochastic process is formulated. An established method for generating constrained surrogates conforming to the properties of a linear Gaussian process is the Iterative Amplitude Adjusted Fourier Transform (iAAFT), which has become quite popular (Teolis, 1998; Schreiber and Schmitz, 1996, 2000; Kugiumtzis, 1999). This type of surrogate time series retains the signal distribution and amplitude spectrum of the original time series, and takes into account a possibly nonlinear and static observation function due to the measurement process. The method uses a fixed point iteration algorithm for achieving this, for the details of which we refer to Schreiber and Schmitz, 1996, 2000.

Wavelet-based surrogate generation is a fairly new method of constructing surrogate for hypothesis testing of nonlinearity which applies a wavelet decomposition of the time series. The main difference between Fourier transform and wavelet transform is that the former is only localized in frequency, whereas the latter is localized both in time and frequency. The idea of a wavelet representation is an orthogonal decomposition across a hierarchy of temporal and spatial scales by a set of wavelet and scaling functions.

The iAAFT-method has recently been refined using a wavelet-based approach, denoted by WiAAFT (Keylock, 2006), that provides for *constrained realizations*[3] of surrogate data that resembles the original data while preserving the local mean and variance as well as the power spectrum and distribution of the original except for randomizing the nonlinear properties of the signal. The WiAAFT-procedure follows the iAAFT-algorithm but uses the Maximal Overlap Discrete Wavelet Transform (MODWT) where the iAAFT-procedure is applied to each set of wavelet detail coefficients $D_j(n)$ over the dyadic scales 2^{j-1} for $j = 1, \cdots, J$, i.e., each set of $D_j(n)$ is considered as a time series of its own. The main difference between iAAFT and wiAAFT algorithms is that the former is designed to produce constrained, linear realisations of a process that can be compared with the original time series on some measure, while the later algorithm restricts the possible class of realisations to those that retain some aspect of the local mean and variance of the original time series (Keylock, 2008).

Statistical analysis by the concept of *surrogate data* tests for a difference between a *test statistic* computed for the original and linearized versions of the data, i.e., an ensemble of realizations of the null hypothesis linear dynamics. For statistical testing of the null hypothesis of

---

[2] The first three sections of this overview do not differ from those presented in Addo et al. (2012a).

[3] These are surrogate realizations that are generated from the original data to conform to certain properties of the original data, e.g., their linear properties, i.e., mean, standard deviation, distribution, power spectrum and autocorrelation function (Schreiber and Schmitz, 1996, 2000).

3

linearity, we follow Theiler et al. (1992) by using a non-parametric rank-order test. The degree of difference between the original and surrogate data is given by the ranked position of the data asymmetry with respect to the surrogates. For a right-tailed test, we generate at least $N_s = \frac{1}{\alpha} - 1$ surrogates, where α is the level of significance and N_s denotes the number of surrogates. The rank-threshold (or critical value) for right-tailed rank-order test is given by $(1 - \alpha)(N_s + 1)$. The null of linearity is rejected as soon as the rank-order statistic is greater than the rank-threshold. To achieve a minimal significance requirement of 95% ($\alpha = 0.05$), we need at least 19 surrogates time series for right-tailed tests. Increasing the number of surrogates can increase the discrimination power (Schreiber and Schmitz, 1996, 2000; Theiler et al., 1992). The concept of *surrogate data* will be incorporated into the *Delay Vector Variance* method *(below)* to examine the dynamics of an underlying economic indicator.

2.2. *Optimal Embedding Parameters*

In the context of signal processing, an established method for visualising an attractor of an underlying nonlinear dynamical signal is by means of time delay embedding (Hegger et al., 1999). By time-delay embedding, the original time series $\{x_k\}$ is represented in the so-called 'phase space' by a set of delay vectors[4] (DVs) of a given embedding dimension, m, and time lag, τ : $x(k) = [x_{k-\tau}, \cdots, x_{k-m\tau}]$. Gautama et al., 2003 proposed a differential entropy based method for determining the optimal embedding parameters of a signal. The main advantage of this method is that a single measure is simultaneously used for optimising both the embedding dimension and time lag. We provide below an overview of the procedure:

The "Entropy Ratio" is defined as

$$R_{ent}(m, \tau) = I(m, \tau) + \frac{m \ln N}{N}, \tag{1}$$

where N is the number of delay vectors, which is kept constant for all values of m and τ under consideration,

$$I(m, \tau) = \frac{H(x, m, \tau)}{\langle H(x_{s,i}, m, \tau) \rangle_i} \tag{2}$$

where x is the signal, $x_{s,i}$ $i = 1, \cdots, T_s$ surrogates of the signal x, $\langle \cdot \rangle_i$ denotes the average over i, $H(x, m, \tau)$ denotes the differential entropies estimated for time delay embedded versions of a time series, x, which an inverse measure of the *structure* in the phase space. Gautama et al., 2003 proposed to use the Kozachenko-Leonenko (K-L) estimate (Leonenko and Kozachenko, 1987) of the differential entropy given by

$$H(x) = \sum_{j=1}^{T} \ln(T\rho_j) + \ln 2 + C_E \tag{3}$$

where T is the number of samples in the data set, ρ_j is the Euclidean distance of the j-th delay vector to its nearest neighbour, and $C_E(\approx 0.5772)$ is the Euler constant. This ratio criterion requires a time series to display a clear structure in the phase space. Thus, for time series with no

---

[4]Time delay embedding is an established method for visualising an attractor of the underlying nonlinear dynamical signal, when processing signals with structure (Hegger et al., 1999).

4

clear structure, the method will not yield a clear minimum, and a different approach needs to be adopted, possibly one that does not rely on a phase space representation. When this method is applied directly to a time series exhibiting strong serial correlations, it yields embedding parameters which have a preference for $\tau_{opt} = 1$. In order to ensure robustness of this method to the dimensionality and serial correlations of a time series, Gautama et al., 2003 suggested to use the iAAFT method for surrogate generation since it retains within the surrogate both signal distribution and approximately the autocorrelation structure of the original signal. In this article, we opt to use wavelet-based surrogate generation method, WiAAFT by in Keylock, 2006, for reasons already discussed in the previous section.

2.3. 'Delay Vector Variance' method

The Characterisation of signal nonlinearities, which emerged in physics in the mid-1990s, have been successfully applied in predicting survival in heart failure cases and also adopted in practical engineering applications (Ho et al., 1997; Chambers and Mandic, 2001). The 'delay vector variance' (DVV) method (Gautama et al., 2004a) is a recently proposed phase space based method for signal characterisation. It is more suitable for signal processing application because it examines the deterministic[5] nature of a time series and when combined with the concept of surrogate data, provides as additional account of the nonlinear behaviour of the time series. The DVV-analysis is based on the calculation of the *target variance*, σ^{*2}, which is an inverse measure of the predictability of a time series. The algorithm is summarized below:

- For an optimal embedding dimension m and time lag τ, generate delay vector (DV): $x(k) = [x_{k-\tau}, \cdots, x_{k-m\tau}]$ and corresponding target x_k

- The mean μ_d and standard deviation, σ_d, are computed over all pairwise distances between DVs, $\|x(i) - x(j)\|$ for $i \neq j$.

- The sets Ω_k are generated such that $\Omega_k = \{x(i) | \|x(k) - x(i)\| \leq \varrho_d\}$, i.e., sets which consist of all DVs that lie closer to $x(k)$ than a certain distance ϱ_d, taken from the interval $[\min\{0, \mu_d - n_d\sigma_d\}; \mu_d + n_d\sigma_d]$, e.g., uniformly spaced, where n_d is a parameter controlling the span over which to perform the DVV analysis.

- For every set Ω_k, the variance of the corresponding targets, σ_k^2, is computed. The average over all sets Ω_k, normalised by the variance of the time series, σ_x^2, yields the *target variance*, σ^{*2} :

$$\sigma^{*2}(\varrho_d) = \frac{\frac{1}{N} \sum_{k=1}^{N} \sigma_k^2(\varrho_d)}{\sigma_x^2} \tag{4}$$

  where N denotes the total number of sets $\Omega_k(\varrho_d)$

Graphical representation of DVV-analysis is obtained by plotting $\sigma^{*2}(\varrho_d)$ as function of the standardized distance, ϱ_d. The minimum *target variance*, $\sigma_{min}^{*2} = \min_{\varrho_d} [\sigma^{*2}(\varrho_d)]$, which corresponds to the lowest point of the curve, is a measure for the amount of noise which is present

---

[5]This means that the underlying process that generate the data can theoretically be described precisely by a set of linear or nonlinear equations. Thus, the component of a time series that can be predicted from a number of previous samples. Wold (1938)
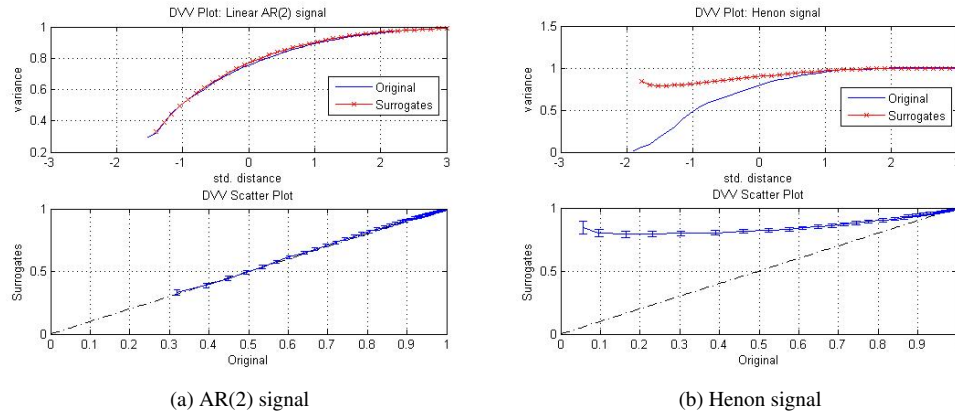
Figure 1: Nonlinear and deterministic nature of signals. The first row of Diagrams 1a and 1b are DVV plots for a linear benchmark signal: AR(2) signal and a nonlinear benchmark signal: Henon signal, where the red line with crosses denotes the DVV plot for the average of 25 WiAAFT-based surrogates while the blue line denotes that for the original signal. The second row of Diagrams 1a and 1b denote the DVV scatter diagrams for those two signals, where error bars denote the standard deviation of the target variances of surrogates.

in the time series. Thus, σ^{*2}_{min} is inversely related to prevalence of the deterministic component over the stochastic one, lowest σ^{*2}_{min} indicating a strong deterministic component. At the extreme right, the DVV plots smoothly converge to unity, as illustrated in Figure 1a and Figure 1b. The reason behind this is that for maximum spans, all DVs belong to the same set, and the variance of the targets is equal to the variance of the time series.

The analysis addressing the linear or nonlinear nature of the original time series is examined by performing DVV analysis on both the original and a set of WiAAFT surrogate time series. Due to the standardisation of the distance axis, these plots can be conveniently combined within a scatter diagram, where the horizontal axis corresponds to the DVV plot of the original time series, and the vertical to that of the surrogate time series. If the surrogate time series yield DVV plots similar to that of the original time series, as illustrated by the first row of Figure 1a, the DVV scatter diagram coincides with the bisector line, and the original time series is judged to be linear, as shown in second row of Figure 1a. If not, as illustrated by first row of Figure 1b, the DVV scatter diagram will deviate from the bisector line and the original time series is judged to be nonlinear, as depicted in the second row of Figure 1b. Statistical testing of the null of linearity using a non-parametric rank-order test (Theiler et al. (1992)) is performed to enhance robust conclusion of results obtained via the DVV-analysis. We refer the reader to Appendix A.1 for more on DVV analysis of some simulated process.

2.4. Data Analysis based on recurrence plots

The method of recurrence plots (RP) was introduced to visualise the dynamics of phase space trajectories (Eckmann et al. (1987)). It is a graphical technique that depicts the different occasions when a dynamical system visits roughly the same area in the phase space. From Takens' embedding theorem (Takens (1981)), the dynamics can be appropriately presented by a recon-

6

struction of the phase space trajectory $\vec{x}(t) = \vec{x}_i \in \mathbb{R}^m$ ($i = 1, \cdots, \eta$, $t = i\Delta t$, where Δt is the sampling rate) in the $m-$dimensional phase space. For a given one-dimensional financial time series $\{u_i\}_{i=1}^N$, the phase space vectors $\vec{x}$ can be reconstructed by embedding the series using Takens' time delay method $\vec{x}_i = (u_i, u_{i+\tau}, \cdots, u_{i+(m-1)\tau})$. The coordinates of this vector correspond to the present and lead values of the series. The parameters m and τ are referred to as the embedding dimension and time delay respectively. We refer to the case for which $m = 1$ and $\tau = 1$ as an unembedded time series.

The recurrence plot is the calculation of an $\eta \times \eta$ matrix

$$R_{i,j}(\varepsilon) = \begin{cases} 1 & : \|\vec{x}_i - \vec{x}_j\| < \varepsilon \\ 0 & : otherwise \end{cases} \quad \vec{x}_i \in \mathbb{R}^m, i, j = 1, \cdots, \eta, \quad \eta = N - (m - 1)\tau, \tag{5}$$

where $\|\cdot\|$ is a norm (e.g Euclidean or maximum norm) and ε is the cut-off distance which defines a region centered at $\vec{x}_i$. If $\vec{x}_j$ falls within this region, the state will be near to $\vec{x}_i$ and is taken to be a recurrence of the state $\vec{x}_i$, which implies $R_{i,j} = 1$. The recurrence plot is square matrix plot of the binary values $R_{i,j}$, in which the matrix element correspond to those calender times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of calender times) (Marwan et al. (2007)). In literature, further variations[6] of the recurrence plots have been proposed for different purposes. In this paper, we make use of a special type of recurrence plot referred to as unthresholded recurrence plots (Iwanski and Bradley (1998)). This recurrence plot is obtained by plotting a matrix of distances $D_{i,j} = \|\vec{x}_i - \vec{x}_j\|$ between the vectors $\vec{x}_i$ and $\vec{x}_j$. As such, it is appropriate to term the unthresholded recurrence plot as *distance plot*. In the section that follows, we will often use the term *distance plot* and recurrence plot interchangeably. In comparing the dynamics of any two time series simultaneously embedded in the same phase space, we employed the cross recurrence plot (Zbilut et al. (1998); Marwan and Kurths (2002)). This is useful in studying the simultaneous occurrence of a state in both series. The cross recurrence plot entails testing for closeness of each point of the first trajectory $\vec{x}_i$ ($i = 1, \cdots, \eta$) with each point of the second trajectory $\vec{y}_i$ ($i = 1, \cdots, \vartheta$) resulting in $\varepsilon \times \vartheta$ array

$$CR_{i,j}(\varepsilon) = \begin{cases} 1 & : \|\vec{x}_i - \vec{y}_j\| < \varepsilon. \\ 0 & : otherwise \end{cases} \tag{6}$$

We provide below a summary of our methodology which can be characterized in two stages:

1. Stage One: Detection of Nonlinearity in the underlying time series.
   (a) Obtain embedded parameters using differential entropy method with wiAAFT surrogates.
   (b) Perform Delay vector variance analysis (DVV) on the series using iAAFT surrogates to detect nonlinearity.
   (c) Statistical testing is perform to validate the results obtained from DVV analysis.
2. Stage Two: Detection and explaining the financial cycle.
   Data analysis is done based on recurrence plots.
   - Underlying financial time series is not embedded before recurrence analysis.
   - The series is embedded using the embedding parameter values obtained from the previous stage.

---

[6]For more details on these variations, we refer the reader to http://www.recurrence-plot.tk/variations.php.

7

Remark 1. *[On anticipating crisis] The distance plot, which is obtained by plotting a matrix of distances $D_{i,j}$ in section 2.4, uses mainly the original data. In other words, the rare events or crisis are detected based on the data available and the detections dates or periods does not change for any sample period. One implication is that it will be a challenge to use it for forecasting financial crisis directly since additional data will be required. We hereby provide below our proposed procedure to enable forecasting crisis based on distance plots:*

1. *Use distance plot to detect and explain crisis on any time series (X_t).*
2. *Model the original time series X_t with some modelling approach. The results from the DVV analysis will facilitate the selection of appropriate class of models to use: linear or nonlinear model.*
3. *from step 2, perform forecast with the model to obtain data points X_{t+h}, $h = 1, 2, 3, \cdots, n$. Thus, extending the original data by n data points.*
4. *Finally, we perform data analysis using recurrence (distance) plot on the extended original data i.e. $(X_1, X_2, \cdots, X_{t+n})$.*

As such, in order to anticipate future crisis and even the possible impact before they happen, we recommend the above step-by-step procedure. It is important to note that the interpretation of any rare event detected by distance plots performed on the extended original data is conditional on the chosen forecasting model used in step 2 of the above procedure.

3. Data Analysis on S&P 500 and NASDAQ Composite

Identification of financial bubbles and crisis is a topic of major concern since it is important to prevent collapses that can severely impact nations and economies. In this section, we perform analysis to characterize and detect nonlinear schemes for the US financial cycle considering both S&P 500 Index and Nasdaq Composite. Firstly, our analysis deals with the use of the recently proposed 'delay vector variance' (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using wavelet-based surrogates. This enables the characterisation of the financial time series without a priori assumption on the dynamics or statistical properties of the series. Secondly, we exploit the concept of recurrence plots[7] on both series to locate hidden patterns, non-stationarity, and to examine the nature of these plots in events of financial crisis. In particular, we show the usefulness of recurrence plots in the diagnosis and detection of financial bubbles and crisis, which have significantly impacted economic upheavals in the past few decades.

In this paper, we consider the daily adjusted closing price of S&P 500 Index and Nasdaq Composite[8] spanning over the period 2nd January, 1990 to 31st August, 2012. Figure 2 and Figure 3 is the plot of the daily adjusted closing price of S&P 500 Index and Nasdaq Composite respectively for the period: 1990-01-02 to 2012-08-31, implying 5716 observations.

To begin with, we opted for the differential-entropy based method (Gautama et al., 2003) to determine the optimal embedding parameters, i.e., the embedding dimension, *m*, and the time

---

[7]The *distance plots* are generated with the Cross Recurrence Plot Toolbox in Matlab provided by Norbert Marwan upon request: http://tocsy.pik-potsdam.de/CRPtoolbox/

[8]The historial financial data can be downloaded from http://quote.yahoo.com/ using the `get.hist.quote` command contained in the `tseries` package in *R*.
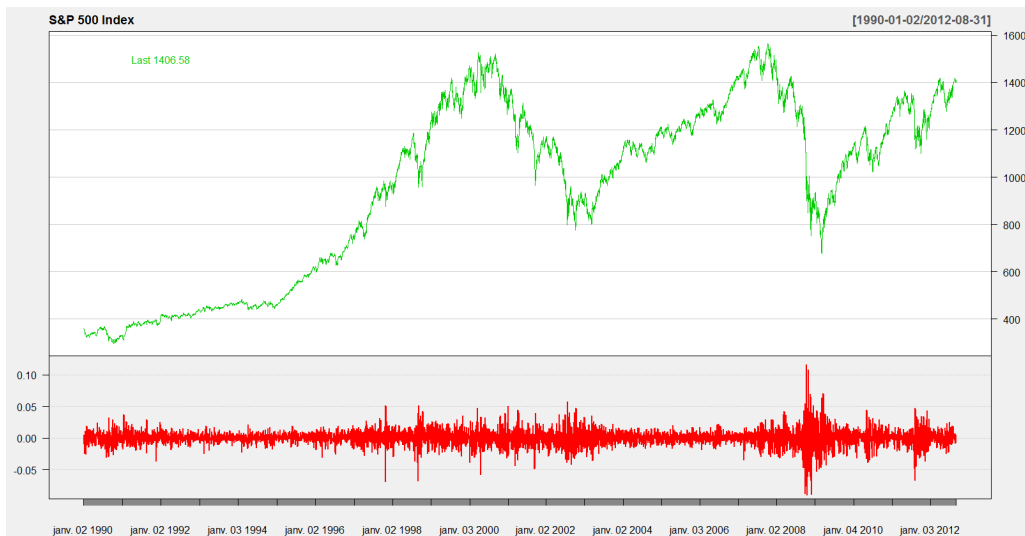
Figure 2: The daily adjusted closing price of S&P 500 Index and it's associated rate of change for the time period 1990-01-02 to 2012-08-31. Thus, the underlying financial time series is of length 5716 observations.
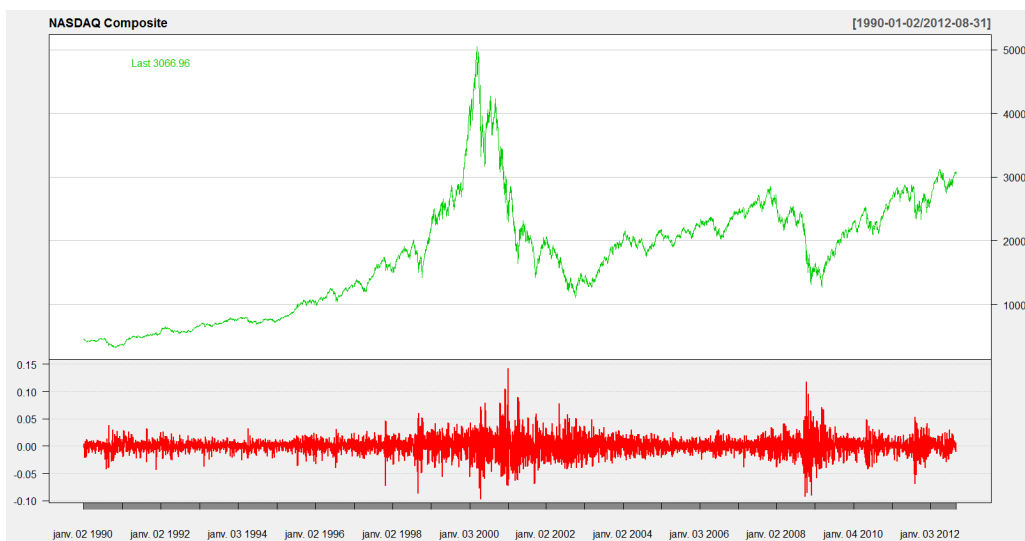

Figure 3: The daily adjusted closing price of Nasdaq Composite and it's associated returns for the time period 1990-01-02 to 2012-08-31. This implies the time series is of length 5716 observations.

lag, τ, for the DVV method with WiAAFT surrogates. The optimal embedding parameters estimated for the S&P 500 in level and it's returns are $(m = 6, \tau = 18)$ and $(m = 2, \tau = 10)$ respectively. This is displayed in Figure 4 as an open circle in the diagram with a clear structure.

9

The embedding parameters obtained for the Nasdaq Composite are ($m = 5, \tau = 19$) and it's associated returns as ($m = 3, \tau = 2$) as showed in Figure 5 . We will use the values of the embedding parameters obtained for the level of S&P 500 and Nasdaq Composite to generate delayed vectors needed to perform the DVV analysis.

(a) Differential-Entropy based method with wiAAFT surrogates on S&P 500 Index. The optimal embedding values are $m = 6$ and $\tau = 18$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.8526$.

(b) Differential-Entropy based method with wiAAFT surrogates on the returns of S&P 500. The optimal embedding values are $m = 2$ and $\tau = 10$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.9948$.
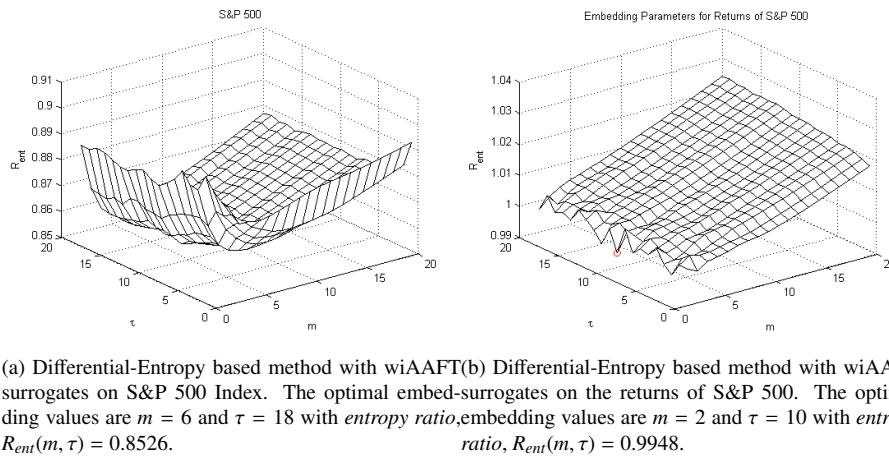
Figure 4: The optimal embedding parameters for S&P 500 Index obtained via the Differential-Entropy based method are indicated as an open circle in the diagrams with a clear structure.

(a) Differential-Entropy based method with wiAAFT surrogates on Nasdaq Composite. The optimal embedding values are $m = 5$ and $\tau = 19$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.8211$.

(b) Differential-Entropy based method with wiAAFT surrogates on the returns of Nasdaq Composite. The optimal embedding values are $m = 3$ and $\tau = 2$ with *entropy ratio*, $R_{ent}(m, \tau) = 0.9959$.
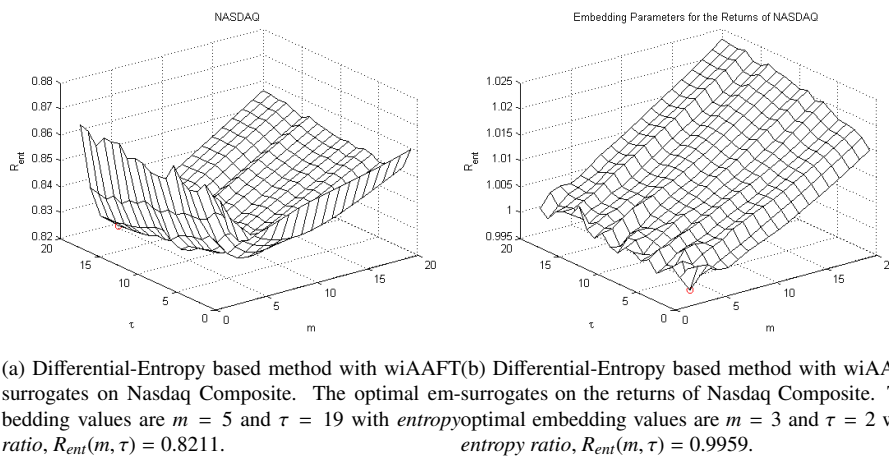
Figure 5: The optimal embedding parameters for Nasdaq Composite obtained via the Differential-Entropy based method are indicated as an open circle in the diagrams with a clear structure.

The results from the DVV analysis with iAAFT surrogates performed on both the S&P 500 and Nasdaq Composite series in level indicates a clear deviation from the bisector on the DVV
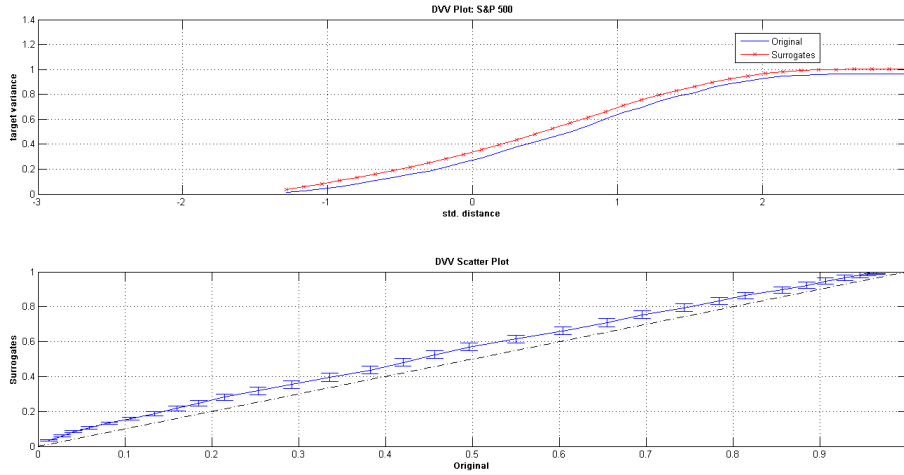
10

Figure 6: This is the DVV analysis with iAAFT surrogates performed on S&P 500 using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, the original time series, S&P 500, exhibits nonlinear dynamics since the surrogates are linear realisations of the original (Schreiber and Schmitz (1996, 2000)).

scatter diagram as shown in Figure 6 and Figure 7 respectively. The DVV plot also shows that the process is neither strictly deterministic or strictly stochastic. Thus, the two original time series exhibits nonlinear dynamics since the iAAFT surrogates are linear realisations of the original (Schreiber and Schmitz (1996, 2000)). Statistical testing of the null of linearity using the non-parametric rank-order test, Table 1, indicates that both time series are nonlinear. The DVV analysis suggests that both time series under consideration behaves more of a nonlinearity with neither a strictly deterministic or strictly component.

Data	Code, H	Rank-Order	Rank-Threshold	Decision
S&P 500	1	26	24.7	Nonlinear Dynamics
Nasdaq Composite	1	25	24.7	Nonlinear Dynamics

Table 1: Results of the non-parametric rank-order test. The null of linearity is rejected as soon as the *Rank-Order* is greater than *Rank-Threshold*. The code H takes the value 0 or 1, where $H = 0$ corresponds to failure of rejecting the null of linearity and $H = 1$ the rejection of linearity for nonlinearity. The number of iAAFT surrogates considered for the DVV-analysis is 25, which is greater than the minimum requirement of 19 surrogates for testing at $\alpha = 0.05$ level of significance.

In the following step, we consider two cases in the identification and characterisation of financial crisis using recurrence plots. The first case consist of perform data analysis with recurrence
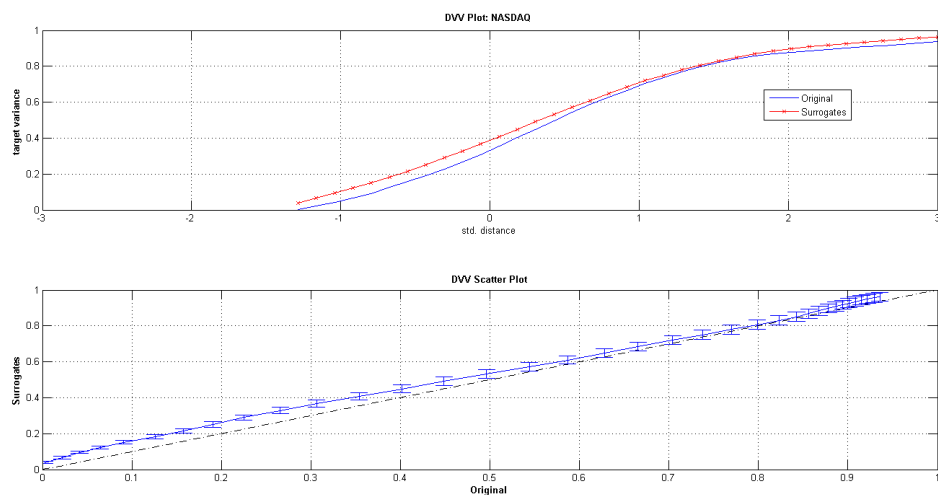
11

Figure 7: This is the DVV analysis with iAAFT surrogates performed on Nasdaq Composite using the embedding parameters obtained via the differential entropy-based method. We clearly observe a deviation from the bisector on the DVV scatter diagram. The DVV plot also indicates that the process is neither strictly deterministic or strictly stochastic. Thus, Nasdaq Composite exhibits nonlinear dynamics since the surrogates are linear realisations of the original.

plots without embedding the underlying financial time series. In the second case, recurrence plots are performed on embedded versions of the underlying financial times using the previously estimated values of embedding parameters.

We provide in Figure 8 and Figure 9 the unthresholded recurrence plots which is sometimes termed *distance plot* (Iwanski and Bradley (1998)) on both underlying financial time series. It is a matrix plot with both vertical and horizontal axes corresponds to calender dates. The colormap corresponds to the distance to the next recurrence of a state in the time series. The results obtained for unembedded and embedded version of both time series do not differ in terms of identification of the financial crisis. The recurrence plot shows the distances between states and enhance understanding of the phase space trajectory of the series. The existence of *butterfly-like* structures along the main diagonal (bisector) indicates a financial crisis. In this case, the start and end of a declining financial activity corresponds the start and end on the formation of *butterfly-like* structure along the main diagonal. In this paper, we designate any *butterfly-like* structure with a minimum size[9] of six months as a financial crisis. A sequential increase in the size of the *butterfly-like* structures along the main diagonal is an indication of a boom in a bubble cycle. In this way, the start of such a sequential increase in size of *butterfly-like* structures can be interpreted as the start of a financial bubble. For instance, Figure 9 shows that the Nasdaq Composite's dot-com bubble started from late 1995 and reached it's peak in March, 2000. We obtain that the financial crisis that resulted from the collapse of the dot-coms bubble began in

---

[9]The size refers to the length in time from the start and end on the formation of *butterfly-like* structure along the main diagonal
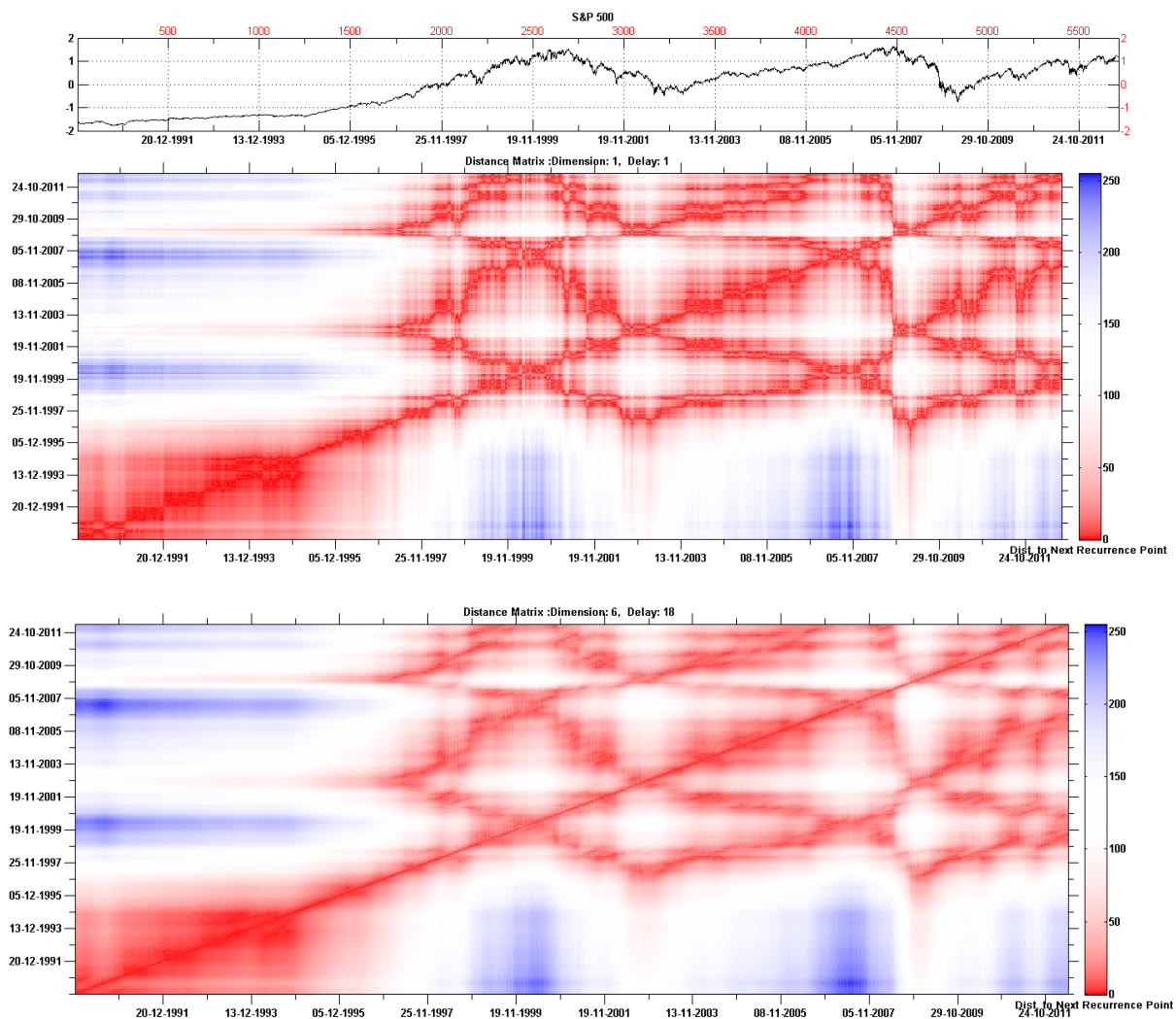
12

Figure 8: This is an unthresholded recurrence plot which is sometimes termed *distance plot* (Iwanski and Bradley (1998)). It is a matrix plot with calender time on both the vertical and horizontal axes for the time period 1990-01-02 to 2012-08-31. The colormap corresponds to the distance to the next recurrence of a state in the time series. This representation shows the distances between states and enhance understanding of the phase space trajectory of S&P 500 and embedded version of S&P 500. The existence of *butterfly-like* structures, of a minimum size of six months, along the main diagonal (bisector) indicates a financial crisis. In this case, the start and end of a financial crisis corresponds the start and end on the formation of *butterfly-like* structure along the main diagonal.

March, 2000 and ended in April, 2003. The 2007-2012 global financial crisis, also known as the Global Financial Crisis and 2008 financial crisis, is also well detected on the recurrence

plot. We discover a financial crisis[10] that began in July 2007 ended in March 2009. This result is interesting since it confirms that stock prices series tend to shift direction in advance of the business cycle. We detect *butterfly-like* structure which corresponds to the existence of financial crisis for the period January, 2011 - January, 2012.

The recurrence plots, displayed in this work, allows to study the recurrence of a state at a particular calender date. Fixing a period or date on the horizontal axis, we are able to observe the recurrence of such events along the calender dates on the vertical axis. We also find that Nasdaq was well exposed and hit by the dot-com bubble of the 1990s compared to S&P 500. Nasdaq, on which many dot-coms traded, rose to record highs as indicated by larger distance to next recurrence. Similar features displayed in Figure 8 and Figure 9 suggests strong financial interdependence.

On studying that the simultaneous occurrence of financial bubbles and crisis in both S&P 500 and Nasdaq Composite, we perform a cross recurrence plot on the series without embedding. The results, as seen in Figure 10, shows that both financial time series were hit by the dot-com bubble.

The returns cycle[11] is often subject to very short-term fluctuations mainly due to transitory events making the peaks of such cycles extremely difficult to date. In the recurrence plots, Figure 11, we are enable to visualise and date these peaks in both returns of the underlying series. The peaks dates are recognised by visualising higher distances to the next recurrence.

Finally, to illustrate our proposed procedure in identifying future crisis and even the possible impact before they happen, as discussed in remark 1, we consider the monthly S&P 500 Stock Price Index[12] from 1990-01-01 to 2012-08-01. It is worth pointing out that the *distance plot* obtained using the monthly data does not differ from Figure 8. The results from the DVV analysis for S&P 500 in Figure 6 suggests that the stock price index is characterised by nonlinear dynamics. In this work, we perform a 12 month ahead forecast of the monthly S&P 500 using an additive nonlinear autoregressive model of the form

$$x_{t+1} = \mu + \sum_{i=1}^{m} g_i(x_{t-(i-1)d}) + \varepsilon_{t+1} \tag{7}$$

where d is time delay, m is the embedding dimension and g_i are nonparametric univariate functions of lagged time series values represented by penalized cubic regression splines (Wood and Augustin (2002); Hastie and Tibshirani (1990)). Model (7) is also referred to as a nonparametric additive autoregressive model (Hastie and Tibshirani (1990)). We assume that error distribution to be gaussian, $\varepsilon_{t+1} \sim i.i.dN(0, \sigma^2_{\varepsilon_{t+1}})$ and apply the model (7) with $m = 2$, $d = 1$, to the monthly S&P 500 Index. The details of estimation of this model is well discussed in Wood and Augustin (2002) and Hastie and Tibshirani (1990). The results of the 12 month ahead forecast based on model (7) are reported in Table 2. Thus, the original data can now be extended by 12 data points as already presented in remark (1). The *distance plot* on the extended data with range 1990-01-01 to 2013-08-01 is displayed in Figure 12. The last *butterfly-like* structure observed in Figure 12 started in March 2012 and ended October 2012. Based on the forecast values obtained from the

---

[10]According to the U.S. National Bureau of Economic Research (NBER), the recession began in December 2007 ended in June 2009,

[11]This cycle is equivalent to the growth rate cycle used in analyzing economic cycles. Suppose, we denote the financial time series by X_t, then the returns is defined by $Y_t = logX_t - logX_{t-1}$.

[12]The data is downloaded from Federal Reserve Economic Data (FRED). The S&P 500 is regarded as a gauge of the large cap U.S. equities market.
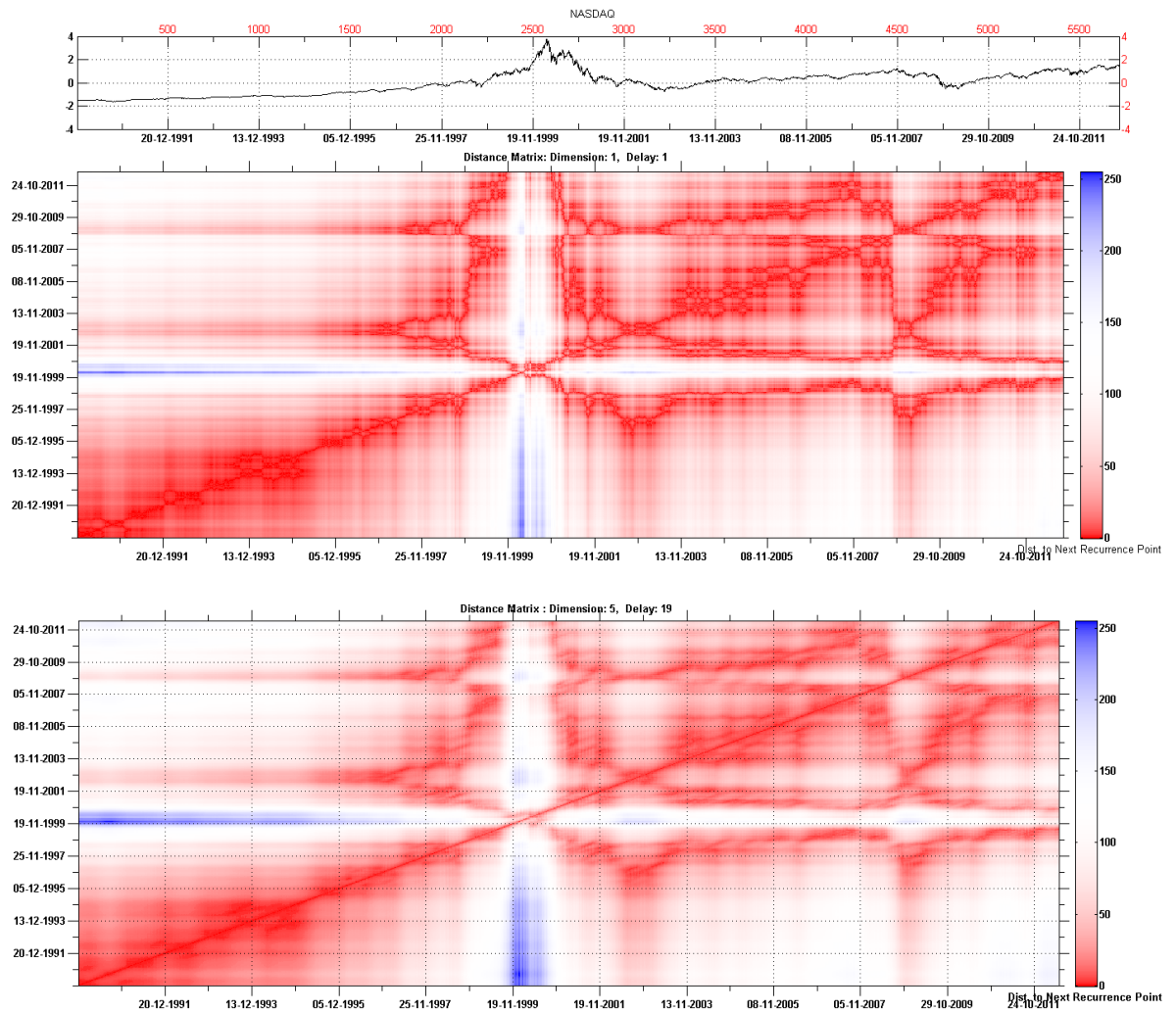
14

Figure 9: The *distance plot* of the Nasdaq Composite and it's embedded version. It is observed that the Nasdaq was well exposed and hit by the dot-com bubble of the 1990s compared to S&P 500. Nasdaq, on which many dot-coms traded, rose to record highs as indicated by larger distance to next recurrence.

model (7) for the monthly S&P 500 Index and the *distance plot* in Figure 12, we do not anticipate a financial crisis for the United States in about six months from August 2012. In studying the possible impact of a financial crisis after it is identified, we consider the monthly S&P 500 stock price Index from January 1990 to February 2008 since a crisis is identified to have began in July 2007. We perform a three month ahead forecast based on model (7) and then a recurrence plot analysis on the extended data. In Figure 13, the region of the two ellipse drawn on the plot shows that crisis that began in July 2007 is a recurrence of the financial crisis that resulted from the
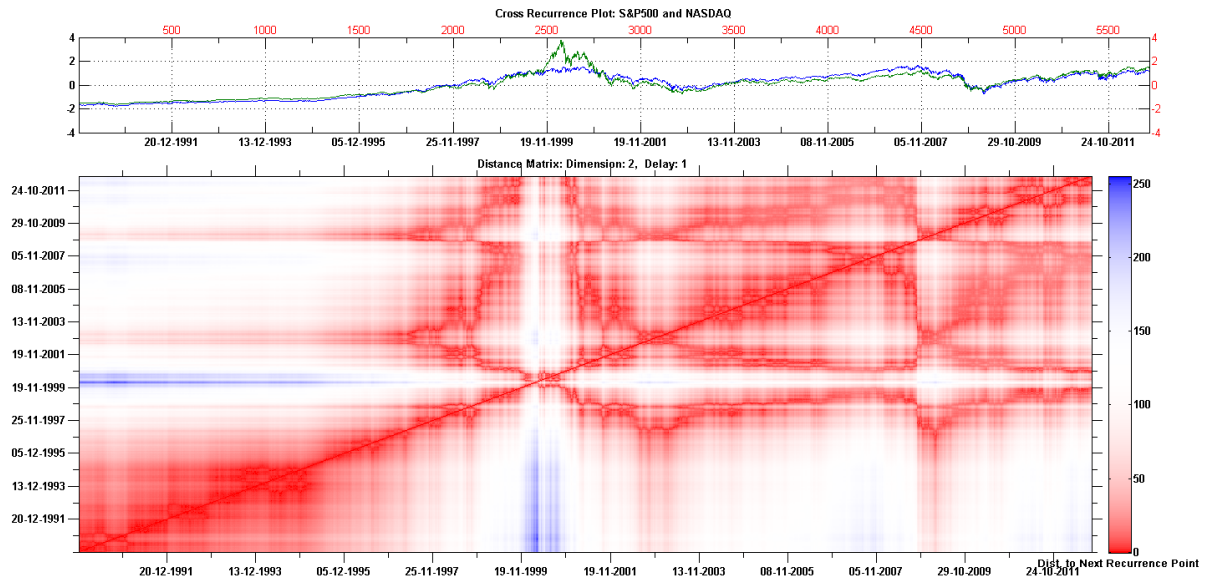
15

Figure 10: This is a cross recurrence plot to visualise the simultaneous occurrence of a similar state in both S&P 500 and Nasdaq Composite. The dot-com bubble clearly occurs in both series.

collapse of the dot-coms bubble. Thus, the possible impact and duration of the financial crisis that began in July 2007 will be similar to the dot-com bubble. We observe from Figure 8, looking at the "wings" of the *butterfly-like* structure, that this results on the possible impact of the crisis that began in July 2007 does not differ much from the dot-coms bubble.

Month	Forecast value
2012:09	1407.371
2012:10	1402.220
2012:11	1395.267
2012:12	1388.223
2013:01	1381.473
2013:02	1375.089
2013:03	1369.070
2013:04	1363.397
2013:05	1358.050
2013:06	1353.009
2013:07	1348.255
2013:08	1343.769

Table 2: This is the 12 month ahead forecast of monthly S&P 500 stock price Index based on the additive nonlinear autoregressive model (7).
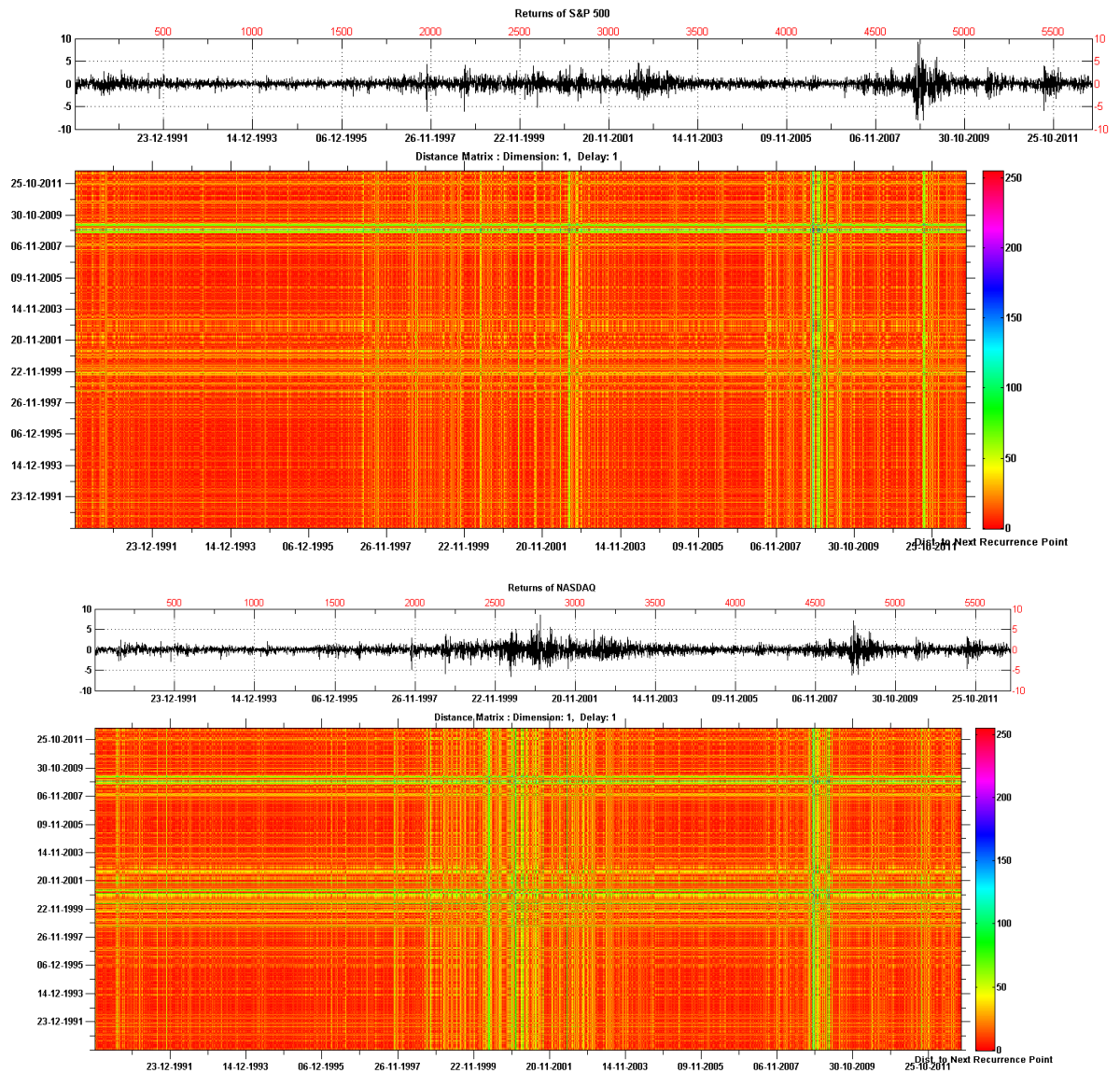
16

Figure 11: The *distance plot* of the growth rate (returns) of S&P 500 and Nasdaq Composite. The growth rate cycle is often subject to very short-term fluctuations mainly due to transitory events making the peaks of such cycles extremely difficult to date. In this case, we are enable to visualise and date these peaks in both underlying series. The peaks dates are recognised by visualising higher distances to the next recurrence.
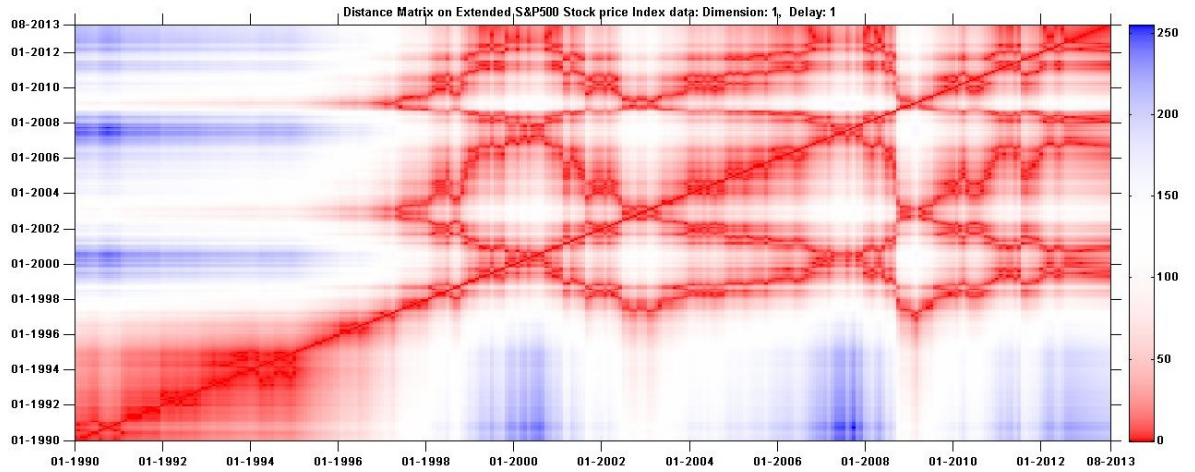
17

Figure 12: The *distance plot* on an extended monthly S&P 500 stock price Index: original data range 1990-01-01 to 2012-08-01 with an additional 12 months ahead forecast obtained via the additive nonlinear autoregressive model (7). There is no signal of a possible occurrence of financial crisis in about six months from August 2012. The last *butterfly-like* structure observed started in March, 2012 and ended October 2012.
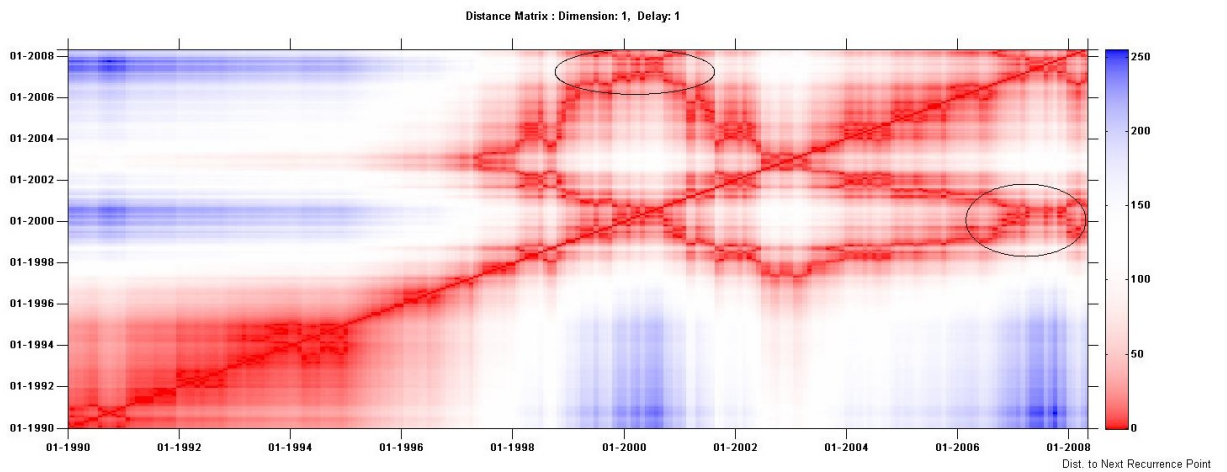

Figure 13: The *distance plot* on an extended monthly S&P 500 price index: original data range 1990-01 to 2008-02 with an additional 3 months ahead forecast obtained via the additive nonlinear autoregressive model (7). The region of the two ellipse drawn on the plot shows that crisis that began in July 2007 is a recurrence of the financial crisis that resulted from the collapse of the dot-coms bubble. Thus, the possible impact and duration of the financial crisis that began in July 2007 will be similar to the dot-com bubble.

18

4. Conclusion

In this paper, we have studied the dynamics of two financial time series using non-parametric methods which are essentially data-driven and carry no *a priori* assumptions on the statistical properties, such as possible non-stationarity, or mathematical structure of the time series. We have provided a comprehensive analysis of the feasibility of our approach as essential in selecting the appropriate class of models suggested by the data itself. Finally, we have demonstrated the usefulness of recurrence plots in identifying, dating and explaining financial bubbles and crisis. The study provides a proposed outline on how to anticipate these rare events and even their impacted before occurrence. The findings from the data analysis with recurrence plots, in this case *distance plots*, shows that these plots are robust to extreme values, non stationarity and to the sample; are replicable and transparent; are adaptive to different time series and finally, can provide better chronology of financial cycles since it avoids revision of crisis dates through time.

19

Appendix A. DVV Plots of simulated Processes

Appendix A.1. DVV Plots of simulated Processes

We provide the structure of the DVV analysis on some simulated processes such as: a Threshold autoregressive process (TAR), linear autoregressive integrated moving average (ARIMA) signal, a Generalised autoregressive conditional heteroskadastic process (GARCH), and a Bilinear process.
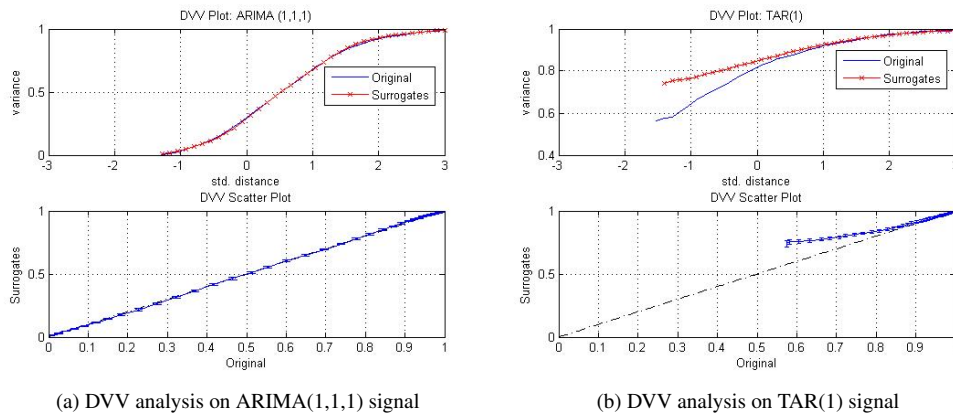

(a) DVV analysis on ARIMA(1,1,1) signal

(b) DVV analysis on TAR(1) signal

Figure A.14: DVV analysis on ARIMA and Threshold Autoregressive signals

(a) DVV analysis on GARCH(1,1) signal
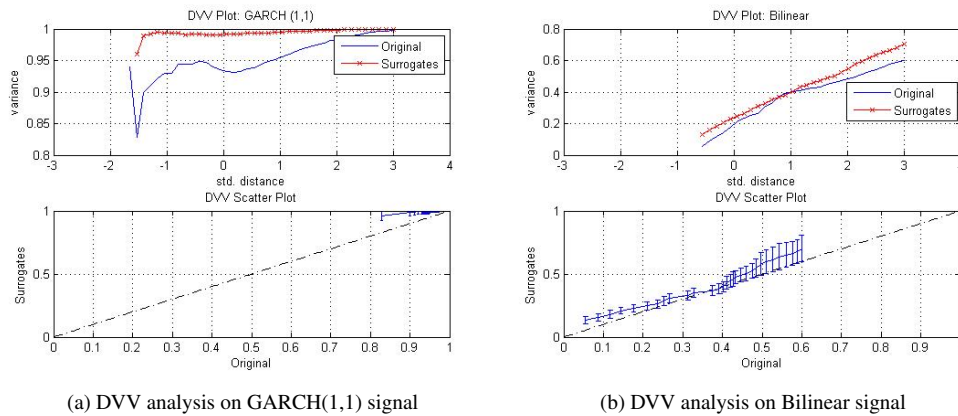
(b) DVV analysis on Bilinear signal

Figure A.15: DVV analysis on GARCH and Bilinear signals

20

References

Addo, P. M., Billio, M., Guégan, D., 2012a. Alternative methodology for turning–point detection in business cycle: A wavelet approach. Documents de travail du Centre d'Economie de la Sorbonne (23).

Addo, P. M., Billio, M., Guégan, D., 2012b. Understanding exchange rate dynamics. in Proceedings of the 20th International Conference on Computational Statistics, Eds. A. Colubi et al, 1–14.

Belaire-Franch, J., 2004. Testing for non linearity in an artificial financial market: a recurrence quantification approach. Journal of Economic Behavior and Organization (54), 483–494.

Chambers, D., Mandic, J., 2001. Recurrent neural networks for prediction: learning algorithms architecture and stability. Chichester, UK: Wiley.

Eckmann, J. P., Kamphorst, S. O., Ruelle, D., 1987. Recurrence plots of dynamical systems. Europhys Lett (5), 973–977.

Gautama, T., Mandic, D. P., Hulle, M. M. V., 2003. A differential entropy based method for determining the optimal embedding parameters of a signal. In Proceedings of ICASSP 2003, Hong Kong IV, 29–32.

Gautama, T., Mandic, D. P., Hulle, M. M. V., 2004a. The delay vector variance method for detecting determinism and nonlinearity in time series. Physica D 190 (3–4), 167–176.

Gautama, T., Mandic, D. P., Hulle, M. M. V., 2004b. A novel method for determining the nature of time series. IEEE Transactions on Biomedical Engineering 51, 728–736.

Hastie, T. J., Tibshirani, R. J., 1990. Generalized Additive Models. Chapman and Hall.

Hegger, R., Kantz, H., Schreiber, T., 1999. Practical implementation of nonlinear time series methods: The tisean package. Chaos 9, 413–435.

Ho, A., Moody, K., Peng, G., Mietus, C., Larson, J., Levy, M., Goldberger, D., 1997. Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. Circulation 96, 842–848.

Holyst, J. A., Zebrowska, M., Urbanowicz, K., 2001. Observation of deterministic chaos in financial time series by recurrence plots, can one control chaotic economy? Eur Phys J B (20), 531–535.

Iwanski, J. S., Bradley, E., 1998. Recurrence plots of experimental data: To embed or not to embed? Chaos 8 (4), 861–871.

Keylock, C. J., 2006. Constrained surrogate time series with preservation of the mean and variance structure. Physical Review E 73, 036707.

Keylock, C. J., 2008. Improved preservation of autocorrelative structure in surrogate data using an initial wavelet step. Nonlinear Processes in Geophysics (15), 435–444.

Kugiumtzis, D., 1999. Test your surrogate data before you test for nonlinearity. Physics Review E 60, 2808–2816.

Kyrtsou, C., Vorlow, C. E., 2005. Complex dynamics in macroeconomics: A noval approach. In: Diebolt, C., Kyrtsou, C. (Eds.). New Trends in Macroeconomics, 223–238.

Leonenko, N. N., Kozachenko, L. F., 1987. Sample estimate of the entropy of a random vector. Problems of Information Transmission 23, 95–101.

Marwan, N., Kurths, J., 2002. Nonlinear analysis of bivariate data with cross recurrence plots. Physics Letters A 302 (5-6), 299–307.

Marwan, N., Romano, M. C., Thiel, M., Kurths, J., 2007. Recurrence plots for the analysis of complex systems. Physics Reports 438 (5–6), 237–329.

Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., Kurths, J., 2002. Recurrence-plot-based measures of complexity and its application to heart-rate variability. Physics Review E (66), 026702.

Schreiber, T., Schmitz, A., 1996. Improved surrogate data for nonlinearity tests. Physics Review Lett. 77, 635–638.

Schreiber, T., Schmitz, A., 2000. Surrogate time series. Physica D 142, 346–382.

Strozzi, F., Gutiérrez, E., Noè, C., Rossi, T., Serati, M., Zaldívar, J. M., 2008. Measuring volatility in the nordic spot electricity market using recurrence quantification analysis. The European Physical Journal-special Topics (164), 105–115.

Takens, F., 1981. Detecting strange attractors in turbulence. Lecture notes in mathematics, 366–387.

Teolis, A., 1998. Computational signal processing with wavelets. Birkhauser.

Theiler, J. D., Eubank, J., Longtin, S., Galdrikian, A., Farmer, B., 1992. Testing for nonlinearity in time series: the method of surrogate data. Physica A 58, 77–94.

Wold, H. O. A., 1938. A study in the analysis of stationary time series. Uppsala, Sweden: Almquist and Wiksell.

Wood, S. N., Augustin, N. H., 2002. Gams with integrated model selection using penalized regression splines and applications to environmental modelling. Ecological Modelling (157), 157–177.

Zbilut, J. P., Giuliani, A., Webber, J. C. L., 1998. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. Physics Letters A (246), 122–128.

Zbilut, J. P., Webber, J. C. L., 1992. Embeddings and delays as derived from quantification of recurrence plots. Physics Letters A (171), 199–203.