Chapter 1
Applications of Methods and Algorithms of Nonlinear Dynamics in Economics and Finance

Abdol S. Soofi, Andreas Galka, Zhe Li, Yuqin Zhang and Xiaofeng Hui

Abstract The traditional financial econometric studies presume the underlying data generating processes (DGP) of the time series observations to be linear and stochastic. These assumptions were taken face value for a long time; however, recent advances in dynamical systems theory and algorithms have enabled researchers to observe complicated dynamics of time series data, and test for validity of these assumptions. These developments include theory of time delay embedding and state space reconstruction of the dynamical system from a scalar time series, methods in detecting chaotic dynamics by computation of invariants such as Lyapunov exponents and correlation dimension, surrogate data analysis as well as the other methods of testing for nonlinearity, and mutual prediction as a method of testing for synchronization of oscillating systems. In this chapter, we will discuss the methods, and review the empirical results of the studies the authors of this chapter have undertaken over the last decade and half. Given the methodological and computational advances

We are grateful to an anonymous referee for helpful comments on an earlier draft of this chapter.

A. S. Soofi (✉)
Department of Economics, University of Wisconsin-Platteville, Platteville, WI 53818, USA
e-mail: soofi@uwplatt.edu

A. Galka
Department of Neuropediatrics, Christian-Albrechts-University of Kiel, Schwanenweg 20, 24105 Kiel, Germany

Z. Li
School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao 066004, Hebei, People's Republic of China

Y. Zhang
School of Public Administration, University of International Business and Economics, Beijing 100029, China

X. Hui
School of Management, Harbin Institute of Technology, Harbin 150001, People's Republic of China

M. Faggini and A. Parziale (eds.), *Complexity in Economics: Cutting Edge Research*, New Economic Windows, DOI: 10.1007/978-3-319-05185-7_1,

of the recent decades, the authors of this chapter have explored the possibility of detecting nonlinear, deterministic dynamics in the data generating processes of the financial time series that were examined. We have conjectured that the presence of nonlinear deterministic dynamics may have been blurred by strong noise in the time series, which could give the appearance of the randomness of the series. Accordingly, by using methods of nonlinear dynamics, we have aimed to tackle a set of lingering problems that the traditional linear, stochastic time series approaches to financial econometrics were unable to address successfully. We believe our methods have successfully addressed some, if not all, such lingering issues. We present our methods and empirical results of many of our studies in this chapter.

Keywords Nonlinear deterministic dynamics · Financial integration · Nonlinear prediction · Synchronization of stock markets · Correlation dimension · Time-delay embedding

Introduction

The traditional empirical financial and economic studies presume the underlying data generating processes (DGP) of the time series observations to be linear and stochastic. However, recent advances in statistical physics, probability theory, and ergodic theory, which are summarized under the rubric of dynamical systems theory and algorithms have enabled researchers to observe complicated dynamics of time series data, and test for validity of these assumptions. These developments include theory of time delay embedding and state space reconstruction of the dynamical system from a scalar time series (Takens 1981; Sauer et al. 1991), methods in detecting chaotic dynamics by computation of invariants such as Lyapunov exponents (Pesin 1977; Wolf et al. 1985) and correlation dimension (Grassberger and Procaccia 1983), surrogate data analysis (Schreiber and Schmitz 1996) and the other methods of testing for nonlinearity (McLeod and Li 1983; Tsay 1986; Brock et al. 1996), and mutual prediction as a method of testing for synchronization of oscillating systems (Fujisaka and Yamada 1983; Afraimovich et al. 1986; Pecora and Carroll 1990).

Traditionally, the numerical algorithms of nonlinear dynamical systems are mostly used in analyses of experimental data of physics and other physical and natural sciences; however, over the last two decades, these methods and algorithms have found extensive use in finance and economics also (Scheinkman and LeBaron 1989; Soofi and Cao 2002a; Soofi and Galka 2003; Das and Das 2007; Zhang et al. 2011; Soofi et al. 2012).

These advances have opened up possibilities of gaining further insights into the dynamics of financial/economic data. Even though from a theoretical point of view these methods are as applicable to economic data as they are to financial data, in practice one observes more frequent applications of these methods to financial data compared to economic data. The reason for this mismatch in applications is low frequency nature of most economic time series data (most economic time series observations are monthly, quarterly, or annual), which leads to limited observations. The algorithms of nonlinear dynamical systems require very large set of time series observations. The financial time series with adequate number of observations for use in nonlinear dynamical analysis could be obtained from the financial markets.

At the outset, we should point out that applicability of these methods and algorithms and the validity of the empirical results hinge on nonlinearity of time series observations. The name nonlinear deterministic dynamics, which is known chaos theory also, should make this requirement absolutely clear. Accordingly, tests for nonlinearity of the series under investigation assume a paramount importance in nonlinear data analyses, and are an absolute requirement before applying any of the above mentioned methods to the data. Nonlinearity is a necessary condition for nonlinear deterministic (chaotic) as well as nonlinear stochastic dynamics.

In this chapter, we will discuss the methods, and review the empirical results of the studies the authors of this chapter have undertaken over the last decade and half. Given the methodological and computational advances of the recent decades, the authors of this chapter have explored the possibility of detecting nonlinear, deterministic dynamics in the data generating processes of the financial time series that were examined. We have conjectured that the presence of nonlinear deterministic dynamics may have been blurred by strong noise in the time series, which could give the appearance of the randomness of the series. Accordingly, by using methods of nonlinear dynamics, we have aimed to tackle a set of lingering problems that the traditional linear, stochastic time series approaches to financial econometrics were unable to address successfully. We believe our methods have successfully addressed some, if not all, such lingering issues. We present our methods and empirical results of many of our studies in this chapter and leave the judgment of how successful we have been in resolving the lingering issues in the financial econometrics to reader.

Specifically, section "Defining Chaotic or Nonlinear Deterministic Dynamics" gives an overview of concepts and definitions of nonlinear dynamical systems. In section "Surrogate Data Analysis and Testing for Nonlinearity", we discuss surrogate data analysis as a test for nonlinearity. Section "Determining Time Delay and Embedding Dimension" reviews time-delay and embedding dimension methods that are used in phase space reconstruction of nonlinear dynamical systems from a single set of observations of the dynamics. In section "Nonlinear Prediction", we discuss the use of nonlinear deterministic method in predictions of the financial time series. Section "Discriminate Statistics for Hypothesis Testing in Surrogate Data Analysis" discusses discriminate statistics that are often used in surrogate data analysis and in tests for detection of chaotic systems. Section "Nonlinear Predictions of Financial Time Series: The Empirical Results" reviews the empirical results of nonlinear prediction of financial time series. In section "Noise Reduction and Increased Prediction Accuracy" the effect of noise reduction on prediction accuracy is examined. Section "Mutual Prediction as a Test for Integration of the Financial Markets" reviews method of mutual prediction as a test for integration of financial markets. Finally, section "Summary and Conclusion" concludes the chapter.

Defining Chaotic or Nonlinear Deterministic Dynamics

It is useful for our subsequent analyses to start with concise definitions of some of the terminologies of nonlinear dynamical systems theory. However, before giving formal definitions of these terms, we give a general description of nonlinear dynamical systems.

Economies (and financial markets), like population biology and statistical physics, consist of large numbers of agents (elements), which are organized into dynamic, volatile, complex, and adaptive systems. These systems are sensitive to the environmental constraints and evolve according to their internal structures that are generated by the relationships among the individual members of the systems. Of course, each of these disciplines has its own peculiarities, the knowledge of which necessitates development of expertise in the respective discipline. However, synthetic microanalytic approach to study the systems is their common characteristic. This implies that one could aim to understand the behavior of the system as a whole by relating the system's behavior to the conducts of its constituent parts on one hand, and by considering interactions among the parts on the other.

For example, in finance one might be interested in learning how trading by thousands of investors in the stock market determines the daily fluctuations in the stock indexes; or in physics, one might be interested to explain how interactions among countless number of atoms result in transformation of a liquid into solid.

Given the evolutionary nature of economic (financial) systems, dynamical systems theory is the method of choice in studying these complex, adaptive systems. A dynamical system is a system whose state evolves over time according to some dynamical laws. The evolution of the system is in accord with working of a *deterministic evolution operator*. The *evolution operator*, which can assume a differential or a difference equation form, a matrix form, or a graph form provides a correspondence between the initial state of the system and a unique state at each subsequent period. In real dynamical systems random events are present, however, in modeling these real systems the random events are neglected.

Let the state of the dynamical system be described by a set of d state variables, such that each state of the system corresponds to a point $\boldsymbol{\xi} \in \mathbf{M}$, where $\mathbf{M}$ is a compact, differentiable d-dimensional manifold. $\mathbf{M}$ is called the *true state space* and d is called the *true state space dimension*.

The states of dynamical systems change over time, hence the state is a function of time, i.e., $\boldsymbol{\xi}(t)$.

In continuous cases a curve or a trajectory depicts the evolutionary path of $\boldsymbol{\xi}(t)$. If the current state of system $\boldsymbol{\xi}(0)$, where one arbitrarily defines the current time $t = 0$, uniquely determines the future states $\boldsymbol{\xi}(t), t > 0$, the system is a *deterministic dynamical system*. If such unique correspondence between the current state and the future states does not exist, the system is called a *stochastic dynamical system*. The completely uncorrelated states are called *white noise*.

In practice, it is not feasible to observe $\boldsymbol{\xi}(t)$, the true states of the dynamical systems. However, measurement of one or several components of the system might be possible. Therefore, using a measurement function $h : \boldsymbol{R}^d \rightarrow \boldsymbol{R}^{d'}$ on the true state $\boldsymbol{\xi}$, we measure a time series $x(t) = h(\boldsymbol{\xi}(t)) + \eta(t)$, where $\eta(t)$ is measurement error (noise) and $d' < d$.

The properties of the evolution operator define the characteristics of the system. A dynamical system is *linear* if its evolution operator is linear; otherwise the system is *nonlinear*.

We need to define *attractor* of a dynamical system before further discussions of the possible forms of behavior of the dynamical systems. To do so, we start with a formal re-statement of *deterministic dynamical systems*.

Start with a system in the initial state of $\boldsymbol{\xi}(0)$. If the system is deterministic, a unique function f^t maps the state at time 0 to state at time t: $\boldsymbol{\xi}(t) = f^t(\boldsymbol{\xi}(0))$. We assume the f^t to be differentiable function, which has a smooth inverse. Such a function is a *diffeomorphism*.

Depending on the structure of f^t, the behavior of $\boldsymbol{\xi}(t)$ for $t \rightarrow \infty$ (after the transient states) varies. In a dissipative dynamical system, where energy of the system is not conserved, all volumes in the state space shrink over time and evolve into a reduced set $\boldsymbol{A}$ called *attractor*. Accordingly, we define an attractor as a set of points in the state space which are invariant to flows of f^t. The transient state is the state in which the process of convergence of the neighboring trajectories to a set of points A of attractor is taking place.

Four types of attractors are observed, which are defined below.

- Fixed points
  The initial state converges into a single point. The time series of such system is given by $x(t) = x(0)$, implying a constant set of observations.
- Limit cycles
  The initial state converges to a set of states, which are visited periodically. The time series corresponding to limit cycles is defined by $x(t) = x(t + T)$, where T is the period of periodicity.
- Limit tori
  A limit torus is the limit cycle with more than one incommensurable frequency in the periodic trajectory.
- Strange attractors
  Strange attractors are characterized by the property of attracting initial states within a certain basin of attraction, while at the same time neighboring initial states on the attractor itself are propagated on the attractor in a way such that their distance will, initially, grow exponentially. When the distance approaches the size of the attractor, this growth will stop due to back-folding effects.
  The time series representing the dynamical systems with strange attractors appear to be stochastic, even though they are completely deterministic. These dynamical systems are called chaotic or nonlinear deterministic dynamics.

We defined nonlinear systems in the context of evolution operators above. However, an intuitive way to gain an understanding of the difference between linear and nonlinear systems is described below.

Perturb the system by x_1 and record its response y_1. Next perturb the system by x_2 and record its response y_2. Then perturb the system by $(x_1 + x_2)$ and record its response y_3. Finally compare $(y_1 + y_2)$ and y_3. If they are equal for any x_1 and x_2 then the system is linear. Otherwise it is nonlinear (Balanov et al. 2009).

Many models depicting chaotic behavior have been developed. Among these models we name the most widely used ones such as the Lorenz attractor (Lorenz 1963), Henon map (Henon 1976), tent map (Devany 1989), and logistic map (May 1976).

Surrogate Data Analysis and Testing for Nonlinearity

As stated above, an extensive literature dealing with different methods for testing for nonlinearity in time series observations has evolved over the last two decades. These methods were used in a number of studies that point to possible nonlinearity in certain financial and economic time series[1] (e.g. Scheinkman and LeBaron 1989; Hsieh 1991; Yang and Brorsen 1993; Kohzadi and Boyd 1995; Soofi and Galka 2003; Zhang et al. 2011; Soofi et al. 2012).

The dynamics of short, noisy financial and economic time series could be the outcome of working of nonlinear determinism in its varieties (periodic, limit tori, and chaotic), stochastic linearity and nonlinearity, and random noise emerging from either or both the dynamics itself and from measurement. Accordingly, in applications of methods and algorithms of nonlinear dynamical systems the first task is to delineate and disentangle all these influences on the observed data set. Given the daunting task of accounting for above listed influences, in practice most analysts focus on determining the role nonlinearity plays in the observed series.

One of the most popular methods of testing for nonlinearity of time series is the surrogate data technique (Theiler et al. 1992). In the surrogate data method of testing for nonlinearity of the series one postulates the null hypothesis that the data are linearly correlated in the temporal domain, but are random otherwise. Among the most popular test statistics for hypothesis testing we mention correlation dimension and some measures of prediction accuracy. We have used both correlation dimension as well as root mean square errors as test statistics for hypothesis testing within the framework of surrogate data analysis on a number of exchange rates and stock market time series studies. We will discuss these quantities below in section "Discriminate Statistics for Hypothesis Testing in Surrogate Data Analysis" after introduction of the method of phase space reconstruction by time-delay embedding.

Presence of noise in the data and insufficient number of observations may point to nonlinearity of a stochastic time series even though the series might be linear (see for example, Osborne and Provencale 1989). To exclude the possibility of receiving such

[1] As it will become clear in the discussion of surrogate analysis below, *nonlinearity* is not a property of a series; it is the absence of the property of linearity that is often detected. However, it is more straightforward, even though less accurate, to speak of presence of nonlinearity in a series throughout this chapter.

misleading signals, surrogate data analysis is often used for testing for nonlinearity of a series. One of the methods used in surrogate data analysis generates a number of surrogates for the original series by preserving all the linear correlations within the original data while destroying any nonlinear structure by randomizing the phases of the Fourier transform of the data. Alternatively one might wish to describe the linear correlations within the original data by generating the linear surrogates from an autoregressive model of order p model, AR(p), and then using the surrogates for estimation of the autocorrelation function (see Galka 2000).

In many practical cases of data analysis, one is faced with a single set of short, noisy, and often non-stationary observations. In such cases, the application of the nonlinear dynamical methods leads to point estimates leaving the analyst without measures of statistical certainty regarding the estimated statistics. One approach to overcome this problem is artificial generation of many time series which by design contain the relevant properties of the original time series, which are obtained through the estimated statistics.

The strategy in surrogate data analysis is to take a contrarian view. The analyst should choose a null hypothesis that contradicts his/her intuition about the nature of the time series under investigation. For example, if one is testing for presence of nonlinear deterministic dynamics in the series, one should select a model that directly contradicts these properties and use a linear, stochastic model to generate the surrogate data, which are different realizations of the hypothesized linear model. Using the surrogate, the quantity of interest, for example, correlation dimension as a discriminating statistics, is estimated for each realization. The next step in this strategy is formation of a distribution using the estimates of the discriminating statistics from the surrogates. The resulting distribution is then used in a statistical test, which might show that the observed data are highly unlikely to have been generated by a linear process.

By estimating the test statistics for both the original series and the surrogates, the null hypothesis that the original time series was linear is tested. If the null is true, then procedure for generating the surrogates will not affect measures of *suspected* nonlinearity. However, if the measure of nonlinearity is significantly changed by the procedure, then the null of linearity of the original series is rejected.

An alternative approach in determining the unknown probability distribution of measures of nonlinearity is the parametric bootstrap method (Efron 1982), which aims to extract explicit parametric models from the data. The validity of this approach hinges on successful extraction of the models from the data. The main shortcoming of parametric bootstrap methods is that one cannot be sure about the true processes underlying the data. The surrogate data method, which can been characterized as a *constrained realization* method, overcomes the weakness of parametric bootstrap method, which can be characterized as a *typical realization* method, by directly imposing the desired structure onto the randomized time series.

To avoid spurious results it is essential that the correct structure (according to the null hypothesis) is imposed on the original series. One approach in ensuring validity of statistical test is determining the most likely linear model that might have generated the data, fitting the model, and then testing for the null hypothesis that the data have been generated by the specified model (Screiber 1999, pp. 42–43).

The number of surrogates to be generated depends on the rate of false rejections of the null hypothesis one is willing to accept (i.e., on the *size* of the test). In most practical applications generating 35 surrogate data series should suffice. A set of values of the discriminating statistics $q^1, q^2, \ldots q^{35}$, is then computed from the surrogates.

Rejection of the null hypothesis may be based either on rank ordering or significance testing. Rank ordering involves deciding whether q^0 of the original series appears as the first or last item in the sorted list of all values of the discriminating statistics $q^0, q^1, q^2, \ldots q^{35}$.

If the qs are fairly normally distributed we may use significance testing. Under this method rejection of the null requires a t value of about 2, at the 95 % confidence level, where t is defined as:

$$t = \frac{|q^0 - \langle q \rangle|}{\sigma_q} \tag{1.1}$$

where $\langle q \rangle$ and σ_q are the mean and standard deviation, respectively, of the series $q^1, q^2, \ldots q^{35}$ (for an in-depth discussion of surrogate data analysis see Kugiumtzis 2002 and Theiler et al. 1992).

Note that a software for generating phase-randomization surrogate data, *fftsurr* (fast Fourier transform surrogates) has been made available by Kaplan (2004); it is written in MATLAB. Phase-randomized surrogate data generated by *fftsurr* have the same spectral density function as the original time series. A further improvement of phase-randomization surrogates can be achieved by creating *improved amplitude-adjusted phase-randomization (IAAPR)* surrogates (sometimes also known as *polished surrogates*). These surrogates have a distribution of amplitudes which is identical to that of the original data, in addition to the preservation of the spectral density function. This is achieved by reordering the original series in a way such that the power spectrum of the surrogates and the original series are (almost) identical.

For data with non-Gaussian distribution, phase-randomized surrogates without amplitude adjustment may result in spurious rejection of the null hypothesis. This result is due to difference between the distributions of the surrogates and the original series. To remedy this problem one should distort the original data so that it is transformed to a series with Gaussian distribution. Then from the distorted original series, now a Gaussian series, a set of surrogates is created by phase-randomization. Finally, the surrogates are transformed back to the same non-Gaussian distribution as the original data (for further details see Galka 2000, Chap. 11).

Soofi and Galka (2003) employed the algorithm of Schreiber and Schmitz (1996) for the generation of IAAPR surrogates in the context of the estimation of the correlation dimension of the dollar/pound and dollar/yen exchange rates. They found evidence of presence of nonlinear structure in the dollar/pound rate, however, no such evidence was found for dollar/yen exchange rate.

Zhang et al. (2011) using the IAAPR algorithm generated 30 surrogate series for 4 daily dollar exchange rates data including Japanese yen, Malaysian ringgit, Thai baht, and British pound for testing for presence of nonlinear structure in the exchange rate series. They found evidence of nonlinear structure in dollar/pound rate. However, it was observed that all the exchange rate series go through periods of linearity and nonlinearity intermittently, a characteristic that was not observed for the simulated data generated from the chaotic Lorenz system.

Testing for nonlinearity of the Chinese stock markets data (Soofi et al. 2012) used algorithms that generate phase-randomization surrogates and amplitude-adjusted surrogates (Kaplan 2004), and found evidence of nonlinearity in all three stock market indices in China: Hong Kong stock Index (HSI), Shanghai Stock Index (SSI), and Shenzhen Stock Index (SZI).

Determining Time Delay and Embedding Dimension

Advances in mathematical theory of time-delay embedding by Takens (1981) and later by Sauer et al. (1991) allow understanding of the dynamics of the nonlinear system through observed time series. These algorithms have had a large number of applications in detecting nonlinear determinism from observed time series, e.g., economic and financial time series (Soofi and Galka 2003; Soofi and Cao 2002a; Cao and Soofi 1999; Bajo-Rubio et al. 1992; Larsen and Lam 1992) .

Given the significance of methods of time-delay embedding and phase space reconstruction in nonlinear dynamical time series analyses, we will discuss these techniques in detail below.

Choosing Optimal Model Dimension

Before a discussion of method of determining the optimal embedding dimension, let us define the *dimension* of a set of points. Geometrically speaking a point has no dimension, a line or a smooth curve has a single dimension, planes and smooth surfaces have two dimensions, and solids are three-dimensional. However, a concise, institutive definition is given by Strogatz (1994, p. 404) who stated that "...the dimension is the minimum number of coordinates needed to describe every point in the set."

Given a scalar time series, $x_1, x_2, \ldots, x_N$, one can make a time-delay reconstruction of the phase-space with the reconstructed vectors:

$$\mathbf{V}_n = (x_n, x_{n-\tau}, \ldots, x_{n-(d-1)\tau}), \tag{1.2}$$

where τ is time-delay, d is embedding dimension, and $n = (d-1)\tau + 1, \ldots, N$.

d represents the dimension of the state space in which to view the dynamics of the underlying system. The time-delay (time lag), τ, represents the time interval between the successively sampled observations used in constructing the d-dimensional embedding vectors.[2]

According to the embedding theorems (Takens 1981; Sauer et al. 1991) if the time series is generated by a deterministic system, then there generically exists a function (a map) $\mathbf{F}: \ R^d \mapsto R^d$ such that

$$\mathbf{V}_{n+1} = \mathbf{F}(\mathbf{V}_n), \tag{1.3}$$

if the observation function of the time series is smooth, has a differentiable inverse, and d is sufficiently large. The mapping has the same dynamic behavior as that of the original unknown system in the sense of topological equivalence.

In practical applications, we usually use a scalar mapping rather than the mapping in (1.3), that is,

$$x_{n+1} = f(\mathbf{V}_n), \tag{1.4}$$

which is equivalent to (1.3).

In reconstructing the phase space, the remaining problem is how to select the τ and d, i.e., time-delay and embedding dimension, in a way that guarantees existence of the above mapping. But in practice, because we have only a finite number of observations with finite measurement precision, a good choice of τ is important in phase space reconstructions. Moreover, determining a good embedding dimension d depends on a judicious choice of τ. The importance of choosing a good time-delay is that it could make minimal embedding dimension possible. This implies that optimal determination of embedding-dimension and time-delay are mutually interdependent.

There are several methods to choose a time delay τ from a scalar time series, such as mutual information (Fraser and Swinney 1986) and autocorrelation function methods.

The more interesting issue is the choice of the embedding dimension from a time series. Generally there are three basic methods used in the literature, which include *computing some invariant (e.g., correlation dimension, Lyapunov exponents) on the attractor* (e.g., Grassberger and Procaccia 1983), *singular value decomposition* (Broomhead and King 1986; Vautard and Ghil 1989), and *the method of false neighbors* (Kennel et al. 1992). However, all these methods contain some subjective parameters or need subjective judgment to choose the embedding dimension.

Dealing with the problem of subjective choice of embedding dimension Cao (1997) modified the method of false neighbors and developed a method of the *averaged false neighbors*, which does not contain any subjective parameter provided the time-delay has been chosen. A more general method based on zero-order approximations has been developed by Cao and Mees (1998), which can be used to determine the embedding dimension from any dimensional time series including scalar and multivariate time series.

[2] For details, see an excellent introductory book by Hilborn (1994).

For an unfolding of a time series into a representative state space of a dynamical system, optimal embedding dimension d and time delay τ are required. The methods of computing embedding dimension and time delay, however, presuppose prior knowledge of one parameter before estimation of the other. Accordingly, calculating one parameter requires exogenous determination of the other.

Soofi et al. (2012) adopted the method of simultaneous estimation of embedding dimensions and time delays.[3] They selected that combination of the embedding dimension and time delay in generation of the dynamics that would lead to the minimum prediction error using nonlinear prediction method.

Specifically, let $\zeta_i = f(d_j, \tau_k, \eta_i)$, $[i = 1, \ldots, N;\ j = k = 1, \ldots, M]$, where ζ_i, d_j, τ_k, and η_i are the ith prediction error, the jth embedding dimension, the kth time delay, and the ith nearest neighbors, respectively. Then one would search for that combination of d_j, τ_k, and η_i that minimizes ζ_i.

Below we briefly describe the Cao method. Note that the method takes τ as given, however, the method estimates an embedding dimension that minimizes the prediction error.

For a given dimension d, we can get a series of delay vectors $\mathbf{V}_n$ defined in (1.2). For each $\mathbf{V}_n$ we find its nearest neighbor $\mathbf{V}_{\eta(n)}$, i.e.,

$$\mathbf{V}_{\eta(n)} = \operatorname{argmin}\{||\mathbf{V}_n - \mathbf{V}_j|| : \quad j = (d-1)\tau + 1, \ldots, N, j \neq n\} \tag{1.5}$$

Note, $\eta(n)$ is an integer such that

$$||\mathbf{V}_{\eta(n)} - \mathbf{V}_n|| = \min\{||\mathbf{V}_n - \mathbf{V}_j|| : \quad j = (d-1)\tau + 1, \ldots, N, j \neq n\}$$

where the norm

$$\begin{aligned} ||\mathbf{V}_n - \mathbf{V}_j|| &= ||(x_n, x_{n-\tau}, \ldots, x_{n-(d-1)\tau}) - (x_j, x_{j-\tau}, \ldots, x_{j-(d-1)\tau})|| \\ &= [\sum_{i=0}^{d-1} (x_{n-i\tau} - x_{j-i\tau})^2]^{1/2}. \end{aligned}$$

Then we define:

$$E(d) = \frac{1}{N - J_0} \sum_{n=J_0}^{N-1} |x_{n+1} - x_{\eta(n)+1}|, \; J_0 = (d-1)\tau + 1. \tag{1.6}$$

where $E(d)$ is the average absolute prediction error of a *zero-order approximation* predictor for a given d. Note that a zero order predictor f is $\hat{x}_{n+1} = f(\mathbf{V}_n)$ and $\hat{x}_{n+1} = x_{\eta(n)+1}$, where $\eta(n)$ is an integer such that $\mathbf{V}_{\eta(n)}$ is the nearest neighbor of $\mathbf{V}_n$. Furthermore, note that the N in (1.6) represents only the number of available

[3] The method was suggested by Liangyue Cao.

data points for fitting, which does not include the data points for out-of-sample forecasting.

To choose the embedding dimension d_e, we simply minimize the E, i.e.,

$$d_e = \text{argmin}\{E(d) : \ d \in \mathrm{Z} \ \text{ and } \ d \geq 1\}. \tag{1.7}$$

The embedding dimension d_e we choose gives the minimum prediction error if we use a zero-order approximation predictor. It is reasonable to infer that this d_e will also give good predictions if we use a high-order (e.g., local-linear) approximation predictor, since a high-order predictor is more efficient than a zero-order predictor when making out-of-sample predictions.

In practical computations, it is certainly impossible to minimize the E over all positive integers. So in real calculations we replace (1.7) with

$$d_e = \text{argmin}\{E(d) : \ 1 \leq d \leq D_{\text{max}}\}, \tag{1.8}$$

where D_{max} is the maximum dimension with which one would like to search the minimum value of $E(d)$.

In summary, the above method is to find the embedding dimension by minimizing the 1-step prediction errors using a zero-order approximation predictive model. For details about this method, see Cao et al. (1998a).

Nonlinear Prediction

Reconstruction of phase space from a scalar time series allows prediction of the series. The reconstructed phase space allows approximation of a function representing the dynamics that could be used for prediction. Below we discuss the local-linear prediction method as one of the methods used in function approximation.

Local-Linear Prediction

Having solved the problem of choosing embedding dimension and time-delay for the vectors $\mathbf{V}_n$ defined in (1.2) we now use model (1.4) for prediction.

The next problem is how to approximate function f. Several approximation techniques, such as local-linear approximation, polynomial approximation, neural networks, radial basis function, and wavelet decomposition are available. One of the more straight forward method is local-linear approximation, because it requires a lower computational time.

Suppose we have N_f samples of time series data available for fitting the function, i.e., we have $x_1, x_2, \ldots, x_{N_f}$.

Therefore we have time-delay vectors $\mathbf{V}_n, \ n = J_0, J_0 + 1, \ldots, N_f$

and $J_0 = (d-1)\tau + 1$. We want to predict x_{N_f+1}.

Steps in the local-linear approximation method are listed below:

1. Impose a metric on the delay-vector space, denoted by $||\ ||$. An example is the root-square norm, i.e, $||\mathbf{a}|| = ||(a_1, a_2, \ldots, a_d)|| = (\sum_{i=1}^{d} a_i^2)^{1/2}$.
2. Find the l nearest neighbors of $\mathbf{V}_{N_f}$, denote them by $\mathbf{V}_{j_1}, \mathbf{V}_{j_2}, \ldots, \mathbf{V}_{j_l}$, $J_0 \le j_k < N_f$, $(k = 1, 2, \ldots, l)$, then for any $k = 1, 2, \ldots, l$, $||\mathbf{V}_{j_k} - \mathbf{V}_{N_f}|| \le ||\mathbf{V}_n - \mathbf{V}_{N_f}||$ ($J_0 \le n < N_f$ and $n \ne j_k$ for any $k = 1, \ldots, l$).
3. Construct a local-linear predictor, regarding each neighbor $\mathbf{V}_{j_k}$ as a point in the domain and x_{j_k+1} as the corresponding point in the range. That is, fitting a linear function to the l pairs $(\mathbf{V}_{j_k},\ x_{j_k+1})$ $(k = 1, 2, \ldots, l)$.

   We use the least-squares method to fit this linear function. Denote it by $\hat{F}$, then we have $\sum_k |x_{j_k+1} - \hat{F}(\mathbf{V}_{j_k})|$ minimized.
4. The predicted value of x_{N_f+1} is $\hat{F}(\mathbf{V}_{N_f})$, i.e.,

$$\hat{x}_{N_f+1} = \hat{F}(\mathbf{V}_{N_f}).$$

Discriminate Statistics for Hypothesis Testing in Surrogate Data Analysis

In this section we will discuss two quantities that we have used in various empirical studies as a discriminating statistics in hypothesis testing for nonlinearity of the financial data that were under consideration.

Correlation Sum and Correlation Dimension

One could select from a set of measures as the test statistics in surrogate data analysis as the first step in determining the behavior of the time series. One of the more popular discriminating statistics in nonlinear dynamical system analysis is correlation dimension. Moreover, in addition to being used as a discriminating statistics in hypothesis testing for presence of nonlinearity in the data, correlation dimension may point to the chaotic nature of the nonlinear dynamical system. This is due to the observation that stochastic processes always use all available dimensions of the state space, while deterministic processes may evolve on a manifold of much lower dimension. This results in the observation that the fractal dimensions are substantially smaller than d-degree of freedom of the dynamical system leading to the evidence of determinism. Below, we give a formal definition of correlation dimension.

Starting with a scalar time series, $x_1, x_2, \ldots, x_N$, which might describe the states of a system or may be the result of a time delay embedding of a univariate time series described by

$$\boldsymbol{x}_i = (x_1, x_{i-\tau}, x_{i-2\tau}, \ldots, x_{i-(d-1)\tau}), \tag{1.9}$$

where τ and d are the time delay and the embedding dimension, respectively.

From these vectors the correlation sum[4] is defined by:

$$C(r) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(r - \|\mathbf{x}_i - \mathbf{x}_j\|), \tag{1.10}$$

where $I(.)$ is an indicator function, such that $I(x) = 1$ for positive x and $I(x) = 0$ otherwise. $\|.\|$ denotes maximum norm, though other norms could also be employed. $C(r)$ estimates the probability of finding two vectors in the set which are separated by a distance not larger than a radius r (in d-dimensional state space).

To avoid spurious results due to unwanted dynamic correlations in the set of vectors $\mathbf{x}_i$ it is advisable to omit all those distances $\|\mathbf{x}_i - \mathbf{x}_j\|$ from the correlation sum for which $\mathbf{x}_i$ and $\mathbf{x}_j$ are too close together in time, i. e. for which $i - j < W$ with a fixed integer parameter W (Theiler 1986). The absence of this correction corresponds to $W = 1$. The choice of W is not critical, provided a sufficiently large value is chosen.

For sufficiently small radius r the correlation sum is expected to display a scaling

$$C(r) = a\, r^{d_c} \tag{1.11}$$

a is a constant. Hence the correlation dimension d_c can be obtained by

$$d_c = \lim_{r\to 0} d_c(r) = \lim_{r\to 0} \frac{\partial \log C(r)}{\partial \log r}. \tag{1.12}$$

The derivative is carried out numerically and yields a dimension estimate $d_c(r, m)$, which still depends on radius r and embedding dimension m.

An Information Theoretic Approach in Estimating a Test Statistics

A number of existing methods for direct testing of nonlinearity such as highly popular residual-based methods, and bispectrum (Hinich 1982) exists. However, none of these methods provide an efficient test statistics that is based on a discrete parametric model. The discrete parametric modeling or information theoretic method of testing for nonlinearity provides such an efficient test statistics (Galka and Ozaki 2001).

Given a time series x_i, $i = 1, \ldots, N$, with zero mean and unit variance (this can be realized by simple linear transformation), we can get an autoregressive model

$$x_i = f(x_{i-1}, \ldots, x_{i-p}) + \eta_i, \tag{1.13}$$

[4] We use the term *correlation sum* because we are dealing with discrete time series. In cases that deal with continuous time series, the term *correlation integral* is used.

where p is the model order and η_i is the dynamical noise. Take $f(\cdot)$ to be a linear function, we get an AR(p) model

$$x_i = \sum_{j=1}^{p} a_j x_{i-j} + \eta_i =: \hat{x}_i + \epsilon_i, \tag{1.14}$$

where $\hat{x}_i$ is the prediction value or conditional mean of x_i.

An exponential autoregressive (ExpAR) model is defined as follows:

$$x_i = \sum_{j=1}^{q} (a_j + b_j \exp(\frac{-x_{i-1}^2}{h}))x_{i-j} + \eta_i =: \hat{x}_i + \eta_i, \tag{1.15}$$

where the bandwidth h for each time series can be estimated by

$$h = -\frac{\max x_{i-1}^2}{\log c}, \tag{1.16}$$

and c is a small number selected in advance. The choice is based on the idea of selection of a bandwidth, h, such that the exponential term becomes essentially zero for large amplitude. Since the exponential function is always positive it never becomes exactly zero. Therefore, we must assign a very small positive number for the exponential term and we call this small constant c. If we choose $\log(c) = -30$, this corresponds to $c = 9.3576e - 014$. This is a number close to the machine precision of computers.

Based on the two models represented in Eqs. (1.2) and (1.3), the test statistic is constructed as follows:

$$\delta(p) := \frac{1}{N}(AIC(\mathrm{AR}(p)) - AIC(\mathrm{ExpAR}(q))) \tag{1.17}$$

where AIC is the Akaike Information Criterion (Akaike 1974) and is defined as

$$AIC = N \log[\frac{1}{N-p} \sum_{i=p+1}^{N} (x_i - \hat{x}_i)^2] + 2(P+1) \tag{1.18}$$

where p is model order, and P denotes the number of the parameters a_i and b_j in the model.

Model (1.17) uses AIC as a measure of the quality of fitted model. Note that the smaller the value of AIC, the better the selected model for the data that is being modeled. Accordingly, $\delta > 0$ implies that the nonlinear *ExpAR(p)* model is a better model compared to the linear *AR(p)* in fitting the data. For $\delta < 0$, the models reverse role.

The Empirical Results Based on the Information Theoretic Model Using the Exchange Rates

Zhang et al. (2011) tested for nonlinearity of the daily dollar exchange rates time series. They used the discrete parametric modeling approach (Galka and Ozaki 2001) to compute an efficient test statistic for nonlinearity of daily dollar exchange rates for 3 Asian currencies and British pound series.

To explore whether the underlying dynamics of Asian financial systems went through changes during the Asian Crisis of 1997–1998, they examined the nonlinear properties of currencies of Thailand and Malaysia before, during, and after Asian financial crises, and obtain highly interesting results. They performed the same analysis using yen and pound rates also. They used yen as the currency of an Asian industrialized country that was immune to the Asian Contagion. They used the pound rate because of observed nonlinear structure in the series by other researchers (Soofi and Galka 2003) as a non-Asian currency for the purpose of a control time series in the study.

According to the results of nonlinearity test, Thai baht shows a nonlinear structure for pre-crisis period. However, the nonlinearity is totally absent for crisis and post crisis periods. For Malaysian ringgit, they observe a mild nonlinearity which corresponds to a period of time close to the early July of 1997, when the monetary authorities in Thailand abandoned the pegging of baht to the dollar this may imply cross-country contagion effect. Again, the data support no nonlinearity of the currency during and after the crisis periods. For Japanese yen, they find no evidence of nonlinearity for any period under study here. Finally, for British pound, they observe a very mild nonlinearity in pre-crisis period, observe no evidence of non-linearity during the crisis period, and detect evidence of a very weak nonlinearity immediately after the first post-crisis period. For the remaining post-crisis periods no evidence of nonlinearity is present. Based on these observations one may conclude that a period of high nonlinearity of the exchange rate may be a prelude to a major financial crisis. Constant monitoring of the behavior of an exchange rate using the present method may be a highly effective early warning system for financial crisis and collapse of currency value.

Nonlinear Predictions of Financial Time Series: The Empirical Results

Soofi and Cao in several works (1999, 2002a) used the nonlinear prediction (local linear approximation) method discussed in section "Nonlinear Prediction" for out-of-sample forecasting of several foreign exchange rates. In all of these prediction exercises the nonlinear prediction method out-performed the competing predictors.

Specifically Cao and Soofi (1999) predicted five daily dollar exchange rates time series: Canadian dollar (Ca$), British pound, German mark, Japanese yen, and French franc, from October 1, 1993 to October 3, 1997. In that study they found evidence that the exchange rate data tested have some deterministic dynamics. In fact, from the theoretical patterns of embedding dimensions for different systems showed that it is very unlikely that the above exchange rate return data are generated by purely random processes. They may be generated by high dimensional systems contaminated by (measurement) noise or nonlinear deterministic systems with stochastic driving forces, i.e., dynamic noise and measurement noise.

Furthermore they tested out-of-sample prediction of the above five exchange rate return time series using the local linear method. They evaluated the prediction by local-linear method with mean value predictor, and calculated the root-mean-square error. The results showed their predictions outperform the mean value predictor for the pound/dollar and the yen/dollar rate returns, but not for the three remaining exchange rate returns.

Soofi and Cao (2002a) used the same prediction method in prediction of monthly black market renminbi/dollar (Feb. 1955–June-1989), monthly black market rial/dollar (Jan. 1957–May 1988), and daily fixed renminbi/dollar (4 Jan. 1993–29 Dec. 2000) exchange rates. They found that in all cases the nonlinear prediction method out-performed the benchmark mean predictor.

Finally, Soofi and Cao (1999) performed out-of-sample predictions on daily peseta/dollar spot exchange rates using a simple nonlinear deterministic technique of local linear predictor. They compared the predictions based on local-linear method with those by two simple benchmark predictors: random walk model and mean-value predictor. The results on the differenced time series indicate that their predictions are better than those by the random walk model, and marginally better than the results from the mean-value predictor.

Noise Reduction and Increased Prediction Accuracy

It is well known that noise can seriously limit the performance of prediction techniques on time series. Effective methods are currently still lacking on noisy time series forecasting. The main difficulty is the absence of prior knowledge on what is noise and what is determinism in real time series, especially when the noise takes part in dynamical evolution of the systems, that is, so-called dynamic noise.

There are obviously two possible approaches to predict noisy time series. One is, ignoring the presence of noise, to fit a predictive model directly from noisy data with the faith on possibility to extract the underlying deterministic dynamics from the noisy data. It seems that the technique of neural networks is helpful in doing such kind of fitting (e.g., Albano et al. (1992)). The other is, filtering the noise beforehand, to fit a predictive model from the filtered or noise-reduced data of the noisy time series with the hope that the noise level in the noisy time series has been reduced. We may need to mention that the latter approach should be more effective than the former one at least in the case of short-term predictions (e.g., see Cao et al. (1998b)).

Suppose a noisy time series $\{x_n\}$ is generated in the following way:

$$x_n = h(\mathbf{y}_n) + \eta_n,$$
$$\mathbf{y}_n = \mathbf{f}(\mathbf{y}_{n-1}), \tag{1.19}$$

where h is a measurement function (observable); η_n is additive noise; $\mathbf{y}_n$ and the iterative equation(s) defined by the function $\mathbf{f}$ are the unknown underlying dynamic variable(s) and dynamic equation(s), respectively.

In the former approach, one should fit a predictive model,

$$x_{n+1} = G(x_n, x_{n-\tau}, \ldots, x_{n-(d-1)\tau}) \tag{1.20}$$

based on the observed time series data using some techniques such as neural networks, where d and τ are the so-called embedding dimension and time-delay, respectively.

Obviously the function G in (1.20) changes the additive noise contained in $x_n, x_{n-\tau}, \ldots, x_{n-(d-1)\tau}$ to dynamic noise. In this sense the predictions must be inaccurate if the noise level is relatively high, as the dynamic noise destroys the determinism of the future dynamic behavior completely.

In the latter approach, on the other hand, one should first obtain the noise-reduced data from the observed noisy time series. Assume the noise-reduced time series having been obtained by some noise reduction method, e.g., local projective and singular value decomposition methods (Grassberger et al. 1993), and denote it by $z_1, z_2, \ldots, z_n, \ldots$. So,

$$x_n = z_n + \varphi_n, \quad n = 1, 2, \ldots, \tag{1.21}$$

where the term φ_n is the noise which was removed by the noise reduction method. The ideal result of noise reduction is $z_n = h(\mathbf{y}_n)$ or $\eta_n = \varphi_n$ for each n, see (1.19) for how the $\{x_n\}$ was generated.

Then a predictive model is fitted based on the noise-reduced time series, that is,

$$z_{n+1} = H(z_n, z_{n-\tau}, \ldots, z_{n-(d-1)\tau}). \tag{1.22}$$

Using this predictive model, the future z_{n+1} can be predicted, and *the value predicted can then be regarded as the predicted value of the future x_{n+1}, i.e, the actual data to be observed*. In fact, the predicted value $\hat{z}_{n+1}$ at the time $n+1$ should be the optimal predicted value of x_{n+1} because the noise term φ_{n+1} can never be predicted, see (1.21). If the noise has been significantly reduced in the noisy time series, then the latter approach is expected to give much better predictions than the former one.

Given that most financial time series contain noise: measurement noise, dynamic noise or both of them together, prediction of financial time series is certainly very challenging. It has attracted much attention on development of methods to improve the predictions. Besides traditional linear methods such as autoregression method, some nonlinear methods have also been applied to forecast financial time series (e.g., Cao et al. 1996; Lisi and Medio 1997; Cao and Soofi 1999). These studies are based on Takens' embedding theorem (Takens 1981). In these applications of nonlinear methods or linear methods, however, the predictive models were generally fitted directly from the original noisy data, i.e., the first approach on prediction of noisy time series mentioned above, see the Eq. (1.20). Not much work has been done using the second approach (see the Eq. (1.22) in prediction of financial time series, although it is expected that the second approach should provide better prediction than the first approach in forecasting of noisy time series as explained earlier.

In Soofi and Cao (2002a) both approaches are applied and compared on predicting two real financial time series- daily mark/dollar exchange rate and monthly U.S. Consumer Price Index(CPI), to see how the noise reduction could improve the predictions.

Nonlinear Noise Reduction

Power spectrum is traditionally used in separating noise with a flat or broad band spectrum from the periodic or quasi-periodic signals with sharp spectral lines. This method, however, has been shown inapplicable in dealing with noise in nonlinear time series, particularly chaotic time series, because the method is unable to differentiate between broad-band spectra from signals of chaotic systems and from signals of purely random noise (Grassberger et al. 1993). Therefore, some newly nonlinear noise reduction methods should be used when dealing with noisy nonlinear time series or noisy chaotic time series; for a review of nonlinear noise reduction methods, see e.g., Kantz and Schreiber (1997) and Ott et al. (1994).

The methods of local projective (LP), singular value decomposition (SVD) (Grassberger et al. 1993), and 'simple' nonlinear noise reduction (SNL) (Schreiber 1993) were adopted by Soofi and Cao (Soofi and Cao 2002a) to reduce the noise in the time series tested in the study.

The LP method rests on the hypothesis that the deterministic part of a noisy time series lies on a low-dimensional manifold in a high-dimensional state space reconstructed by the time-delay embedding, while the effect of noise is to distribute the data in the immediate surroundings of the manifold. The method is designed to identify the low-dimensional manifold and project the time series data onto it. Interested readers are referred to Schreiber (1998) for a detailed description of the method and relevant discussions.

Applying SVD to a time series tends to optimize the signal to noise ratio. In filtering data with SVD, the singular vectors of the covariance matrix of the time series are first computed; then the reconstructed m dimensional vectors are projected to a q dimensional space, where q $(< m)$ is the number of singular values computed (see Grassberger et al. (1993) for details).

The idea of the 'simple' nonlinear noise reduction method is to locally approximate the dynamics of the underlying system. Unlike the LP and the SVD methods, this method does not require to project the system to a lower dimensional system.

Mark-Dollar Exchange Rates

Soofi and Cao (2002a) used daily mark-dollar exchange rate time series, for sample observations for the period from October 1, 1993 to October 3, 1997.

Non-filtered data

Prediction test is first conducted on the time series without filtering. That is, no noise reduction is made on the differenced-log time series of mark/dollar exchange rates. This test was done in our earlier work (Cao and Soofi 1999). The RMSE between the out-of-sample predicted and the actual data was 1.08.

The finding that RMSE in the prediction was greater than 1 implies that the prediction by the local linear method is not better than the prediction by a mean value predictor. This negative result was actually expected because the behavior of exchange rates is so complicated that any deterministic predictions may not lead to better performance than the prediction by a simple mean value predictor. High level of noise in the exchange rate time series is also commonly regarded as one of the reasons for the failure of nonlinear deterministic prediction.

Filtered data

Given that one noise reduction method may work well in some cases, while it may not in the others , three sets of filtered data were generated using the simple nonlinear noise reduction (SNL), local projective (LP), and the SVD methods. The last two methods require a prior projection dimension (q). This q is generally not known for real time series (interested readers may consult the literature in noise reduction, e.g., Grassberger et al. (1993), for the selection of q).

The RMSE for the case with LP method for mark/dollar exchange rate was less than 1, which implies that the prediction by the local linear method was better than that by a mean value predictor. This means that noise reduction improves prediction of the exchange rate time series provided that an appropriate noise reduction method as well as suitable parameter values for the method is used. At this stage, the improvement is not statistically significant based on the statistic provided by Harvey et al. (1997) at a 10 % nominal level; however, the improvement is statistically significant at a 20 % nominal level.

U.S. Consumer Price Index

Monthly US consumer price index (CPI) time series was also used by Soofi and Cao (2002b) for out-of-sample prediction exercises. The reason they chose the CPI time series was that it is believed deterministic dynamics should be stronger in the CPI time series than that in the exchange rate time series. Therefore, nonlinear deterministic techniques should have a better chance to provide good prediction on the CPI time series than on the exchange rate time series.

Following the same procedures as for the exchange rate time series, the results for the CPI time series showed that the RMSE (= 0.87) for the non-filtered data was less than 1, which means that the local linear deterministic prediction is better than the mean value prediction. Comparing with the corresponding results of the exchange rate time series, the much smaller RMSE for the CPI time series indicates that the deterministic dynamics in the CPI time series should be stronger than that in the exchange rate time series as we mentioned earlier.

For all other cases, the predictions with noise reduction are even worse than the prediction without noise reduction. This means that noise reduction may have distorted the deterministic dynamics in the CPI time series, therefore, the prediction on the filtered data becomes even more difficult. This could be often the case given it is not known what is the noise and what is the determinism in a real time series. However, this should not be taken as discouragement to use noise reduction in prediction of real time series. It implies that one should carefully select which noise reduction method as well as its related parameter values should be used for a particular time series, because a noise reduction method may work better in some cases, while it may not in other cases.

Mutual Prediction as a Test for integration of the Financial Markets

Another application of the methods of nonlinear dynamics is in testing for integration of economies and financial markets. There exists a vast literature on the subject of financial integration, which uses terms such as integration, globalization, and interdependence interchangeably. However, none of these terms is given a concise, quantitative definition. Soofi et al. (2012), however, using methods from science of nonlinear dynamical systems provided an exact quantitative definition of financial integration and treated terms such as financial integration and interdependence of financial markets synonymously.

The basis for the quantitative definition is the notion that interdependence of two or more financial markets implies that the observed time series of these systems originate from the *different parts* of the *same* dynamical system. The rational for this argument is that the equity markets are the subsystems of the global economic or financial system. Specifically, presence of dynamical interdependence among the subsystems (the individual equity markets) implies that:

1. The subsystems communicate, that is, they are coupled together and information flows between them (news arrival in the financial markets), and/or
2. They are coupled to a common driver, where in the case of the stock markets the driving force is profit motive.

It should be noted that even for coupled, but otherwise independent dynamical systems, it is possible that their temporal evolutions might become “synchronized" as one adjusts the coupling strength between them, even though their temporal evolution might not be identical.

The study of dynamical interdependence of nonlinear systems, commonly known as synchronization in physics literature, has its origin in the works of Fujisaka and

Yamada (1983), Afraimovich et al. (1986) and Pecora and Carroll (1990). A variety of approaches to synchronization studies, including system-subsystem synchronization, synchronization in unidirectional and bidirectional coupled systems, anti-phase synchronization, partial synchronization, pulse-coupled synchronization, and generalized synchronization have been developed.

Oscillating systems evolve along their attractors. In certain situations where the oscillators are asymmetrically coupled, there may exist a one-to-one mapping between each attractor. In presence of such mapping, it is possible to predict behavior of one system given the attractor of another one.

Dynamical interdependence, as described in Rulkov et al. (1995), which adopts a generalized synchronization approach, implies predictability of the *response* system's behavior by the *driving* system. This is the starting point for testing for interdependence of two systems which assumes existence of function ϕ that projects values from the trajectories of the driving system D space into the trajectories in the response system R space. In practice, however, when the degrees and directions of the coupling between the systems are unknown, one aims to reconstruct the dynamics of the two systems by time-delay embedding method, and then estimates statistics for testing for dynamical interdependence between the reconstructed systems. This is the basis for the mutual prediction method for testing for interdependence of two dynamical systems (Pecora and Carroll 1990; Schiff et al. 1996; Breakspear and Terry 2002), a method used by Soofi et al. (2012).

Mutual prediction is a method for testing for synchronization of completely independent, but coupled oscillating systems. Examples of synchronization of completely independent, yet coupled, oscillating systems from biological and physical realms include synchronized intermittent emissions of light by tens of thousand fireflies to random openings of ion channels in cell membranes, to organ pipes, just to name a few. In short, synchronization is interaction among different systems or subsystems, which at the times before or after synchronization, operate independently from each other. This means that these coupled, different, and independent systems or subsystems adjust the time scales of their oscillations due to the interaction (Balanov et al. 2009).

We search for evidence of coupling between these markets by considering their dynamics that are represented by the following differential equations:

$$\frac{d\mathbf{X}}{dt} = f(\mathbf{X}(\mathbf{t}) + \bar{f}(\mathbf{X})\,\boldsymbol{\xi}_1(t)) \tag{1.23}$$

$$\frac{d\mathbf{Y}}{dt} = g[(\mathbf{Y}(\mathbf{t}) + \bar{g}(\mathbf{Y})\boldsymbol{\xi}_2), h_c(\mathbf{X}(\mathbf{t}) + \bar{f}(\mathbf{X})\boldsymbol{\xi}_1(t), \mathbf{Y}(\mathbf{t}) + \bar{g}(\mathbf{Y})\boldsymbol{\xi}_2)] \tag{1.24}$$

where functions f and g generate local dynamics, function h transmits the influence of $\mathbf{X}(t)$ to $\mathbf{Y}(t)$, and constant c measures the strength of coupling. Moreover, $\boldsymbol{\xi}_1(t)$ and $\boldsymbol{\xi}_2(t)$ are random dynamical noise reflecting random decisions of traders in the two markets. These random terms, not to be confused by the measurement noise of equations (1.25) and (1.26) below, may induce oscillatory dynamics in the model, opening up the possibility that the markets meet the self-sustaining oscillations requirement for synchronization.

Let **X** and **Y** be two potentially coupled dynamical systems with the time series observations of x_i and y_i $(i = 1, \ldots, N)$, respectively. Often, in practice, the state variables are not directly observable, and one has no *a priori* knowledge of their individual dynamics or their dynamical interdependence. Instead, their evolutions are measured by the scalar variables

$$x_i(t) = h(\mathbf{X}(t)) + \eta_1(t) \tag{1.25}$$

$$y_i(t) = k(\mathbf{Y}(t)) + \eta_2(t) \tag{1.26}$$

where h and k are the measurement functions (possibly nonlinear), and η_1 and η_2 are the error terms representing noise in the data.

On many occasions one might have to analyze time series data that have values in a wide range. In such cases one should standardize the series by the following transformations:

$$\hat{x}_i = \frac{x_i - \bar{x}}{\sigma_x} \tag{1.27}$$

$$\hat{y}_i = \frac{y_i - \bar{y}}{\sigma_y} \tag{1.28}$$

where $\bar{x}$, $\bar{y}$, σ_x, and σ_y are the mean and standard deviation of the x_i and y_i series, respectively.

Next using the time-delay embedding of section "Determining Time Delay and Embedding Dimension" we would reconstruct the phase spaces for both **X** and **Y**.

Surrogate data analysis is the method of choice in physics and nonlinear dynamical systems analysis. Hence, the mutual prediction method of test for nonlinear interdependence uses this approach also. See section "Surrogate Data Analysis and Testing for Nonlinearity" above for a discussion of this method.

The algorithm of computing the time delay τ with mutual information technique in Soofi et al. (2012) is Shannon's entropy method, and consists of first constructing a histogram for the probability distribution of the data. For details see Soofi et al. (2012).

Algorithm for Mutual Prediction Method

One starts with a possible functional relationship between **X** and **Y** as

$$\mathbf{Y} \stackrel{?}{=} \phi(\mathbf{X}) \tag{1.29}$$

and aims at empirically verifying existence of the functional relationship ϕ between the two reconstructed systems **X** and **Y**. If such a relationship exists, then two close states in the phase space of the **X** system correspond to two close states in the phase space of the **Y** system.

It does not matter which state variable we choose as autonomous or response variable. For measuring nonlinear interdependence what counts is h_c function, and coupling strength coefficient c. Existence of a continuous, differentiable map ϕ, where in presence of synchronization creates a one-to-one correspondence between the orbits of **X** onto the orbits of **Y** in case of $\mathbf{Y} = \phi(\mathbf{X})$, and maps **Y** onto **X** in case of $\mathbf{X} = \phi(\mathbf{Y})$ is the important consideration.

Select an arbitrary point x_0 in the **X** space. Suppose the nearest neighbor of x_0 has a time index of n_{nnd}. Then if function ϕ exists, that is, if the two systems are coupled, then point y_0 in the **Y** space will have point $y_{n_{nnd}}$ as a close neighbor also. This means that the nearest neighbors of both points x_0 and y_0 share the same time indexes.[5] For example, if the nearest neighbor of point x_0 is a three-dimensional vector with time indexes $(1, 5, 8)$, then the vector that is the nearest neighbor of point y_0 has the same time indexes $(1, 5, 8)$.

In implementing the mutual prediction method of testing for nonlinear interdependence of Chinese stock markets, Soofi et al. (2012) followed the method discussed by Breakspear and Terry (2002) which is a modified, improved version of Schiff et al. (1996) as discussed below.

- Construct in **X** a simplex around an arbitrary selected point $x(t_i)$ in time $t = t_i$ with $2d_1^x$ vertices each consisting of another vector in **X**. d_1^x is the embedding dimension of **X**.
- Choose these embedding vectors (vertices) such that the size of the simplex is minimized.
- Denote the points satisfying the criteria of being a vertex in the minimized simplex as $x_j(t_{ij})$, $j = 1, \ldots, 2d_1^x$. Also denote the time indices of the vertices as t_{ij}, $j = 1, \ldots, 2d_1^x$.
- Use the time indices t_{ij} of $x_j(t_{ij})$ to construct a simplex in the state space **Y** with vertices $y(t_{ij})$, $j = 1, \ldots, 2d_1^x$.
- Take the weighted average of the vertices in $y(t_{ij})$ to locate the vector $y(t_{ij})$ that was predicted by the vector $x(t_i)$

$$y_{pred.}(t_i) = \frac{\sum_{k=1}^{2d_1^x} \omega_{ik} y(t_{ik})}{\sum_{k=1}^{2d_1^x} \omega_{ik}} \tag{1.30}$$

  where the weighting factors ω_{ik}, are determined by the distances of the vertices in **X** from $x(t_i)$, giving

$$\omega_{ik} = (|x(t_{ik}) - x(t_i)|)^{-1}. \tag{1.31}$$

- To calculate the mutual prediction error, take the difference of the predicted vector and the actual vector

[5] Note that we have unfolded the time series into d-dimensional space.

$$\epsilon_{y(x)} = |y_{pred}(t_i) - y(t_i)|. \tag{1.32}$$

- To compare the prediction error $\epsilon_{y(x)}$ with a prediction error based on a randomly selected element of the time series observations calculate

$$\epsilon_{rand} = |y_{rand} - y(t_i)|, \tag{1.33}$$

  where y_{rand} is calculated using the same procedure used in prediction of $y_{pred.}(t_i)$, except that the simplex in **X** is a random combination of points on the orbit **X** weighted with respect to another randomly selected point. This corresponds to the null hypothesis of no interdependence between the markets.
- The normalized predicted y, $\nabla_{y(x)}$, as predicted by x, is calculated by

$$\nabla_{y(x)} = \frac{\langle \epsilon_{y(x)} \rangle_{rms}}{\langle \epsilon_{rand} \rangle_{rms}} \tag{1.34}$$

  where $\langle\rangle_{rms}$ is the root mean square.
  $\nabla_{y(x)} = 1$ implies no interdependence (no synchronization). $\nabla_{y(x)} = 0$ implies complete synchronization.
- Calculate the vertices of simplex in **Y** as above and then iterate them H-step ahead on their respective orbits to obtain the vertices $y(t_{ij} + H)$, $j = 1, \ldots, 2d_i^y$
- Compare the weighted predicted vector $y_{pred.}(t_i + H)$, $j = 1, \ldots, 2_i^y$ to the actual forward iterate $y(t_i + H)$ to obtain future prediction errors.
- Normalize the H-step ahead prediction errors by a vector generated from random vertices in **X** to yield the normalized future prediction error:

$$\nabla^H_{y(x)} = \frac{\langle \epsilon^H_{y(x)} \rangle_{rms}}{\langle \epsilon_{rand} \rangle_{rms}} \tag{1.35}$$

  $\nabla^H_{y(x)} = 1$ implies no interdependence between the systems at H-step prediction.

Note that in presence of generalized synchronization the error grows at a rate determined by the Lyapunov exponents, and is less than one for some time steps into the future.

After generating a number of surrogates,which share the spectral density functions with the original time series use one-step ahead mutual prediction method described above, and conduct H forecasts of the original time series and the surrogate time series separately. If the one-step ahead nonlinear prediction errors of the original series are smaller than those for any of the surrogates, predictions are significant.

A plot of H prediction errors as well as prediction interval for the original and surrogate series based on the above mentioned algorithm would aid in determining nonlinear interdependence of the markets.

The deterministic interdependence is detected if the graph of the cross-prediction errors of the original series is below the graphs of cross-prediction errors for the surrogate sets, but above the lower bound of the 95 % confidence interval.

Synchronization of Chinese Stock Markets

Soofi et al. (2012) considered three Chinese stock markets: Shanghai (SSI), Shenzhen (SZI) and Hong Kong (HSI), as nonlinear dynamical oscillating systems. They further considered two indexes at a time for testing and took $\mathbf{X(t)}$ as the driver system and $\mathbf{Y(t)}$ as the response system. Furthermore, they reconstructed the phase space of each stock market as a dynamical system using time series observations of the daily average stock prices.

We note that synchronization can be bi-directional or unidirectional. In a *forced synchronization* one system influences the second one without being influenced by it. One has bidirectional synchronization where both systems are mutually interacting and influencing each other. Hence, in the forced synchronization case, if $\mathbf{X}$ is not influencing $\mathbf{Y}$, it does not necessarily mean that $\mathbf{Y}$ is not influencing $\mathbf{X}$. (for excellent discussions of synchronization, see Balanov et al. (2009)).

They constructed 19 bivariate surrogate data with the same amplitude distribution, auto correlation function, and cross-spectral density function as the original data. However, non-linear structure contained within and between the surrogate series are destroyed. Thus the surrogate algorithm allows testing of the null hypothesis that the time series are produced by a cross-correlated stochastic system.

The results of Soofi et al. (2012) show that there is nonlinear mutual (bidirectional) predictability between SSI and SZI. Moreover, there exists unidirectional predictability from SSI to HSI and from SZI to HSI. However, the results don't provide statistically significant evidence that Hong Kong market predicts the stock markets in mainland China.

In sum, the study concluded that Shanghai, Shenzhen, and Hong Kong stock market data are nonlinear, and are nonlinearly dependent on each other. This implies that the stock index observations of the three stock markets are originated from different parts of the same dynamical system, and hence the markets are well integrated.

Comparing the Results with the Results Based on a Traditional Linear Method

Comparing the results for synchronization of the Chinese stock markets based on mutual prediction method with the results based on a linear method of testing for integration of financial markets, Soofi et al. (2012) used results from Zhu et al. (2003). Zhu et al. (2003) have used cointegration, fractional cointegration, and Granger causality methods in testing for integration of Chinese stock markets. The tests in Zhu et al. (2003) show no evidence of cointegration (either integrated or fractionally integrated) among the stock markets. They could not find any evidence for presence of causality among the markets either. Hence, the mutual prediction method of testing for interdependence of Chinese stock markets data shows completely different results from those obtained by the traditional linear stochastic methods used in Zhu et al. (2003) study. The evidence pointing to nonlinearity of the stock markets as dynamical systems, should support the conclusion that the linear models have failed to detect interdependence, while the mutual prediction method succeeded in finding the evidence of dynamical interdependence between the markets.

Summary and Conclusion

Advances in nonlinear dynamical system theories and methods have opened up new possibilities for applying them in finance and economics. The authors of the present chapter have applied a number of these methods in testing for nonlinearity, predictions, and calculation of invariants such as correlation dimension of the some exchange rate data. They have used these methods in testing for synchronization (interdependence) of the stock markets also.

Even though tests uniformly show presence of nonlinearity in many financial data that were analyzed, determination of whether the data generating processes are deterministic is inconclusive because of the short sample observations and presence of noise in the observed data. Further advances in theory of nonlinear stochastic dynamical systems in the last decades promises to be useful in further applications on the financial data. Applications of these methods, specially mutual prediction method as warning system for imminent emergence of financial contagion is very promising also. Hitherto, the methods of nonlinear dynamic systems unravel the dynamics in many financial time series observations that could not be detected by the tradition linear stochastic methods.

References

Afraimovich, V., Verichev, N., Rabinovich, M.: Stochastic synchronization of oscillation in dissipative systems. Radiophys. Quantum Electron. **29**, 795–803 (1986)

Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Control **19**, 716–723 (1974)

Albano, A.M., Passamante, A., Hediger, T., Farrell, M.E.: Using neural nets to look for chaos. Physica D **58**, 1–9 (1992)

Bajo-Rubio, O., Fernndez-Rodrguez, F.: Chaotic behaviour in exchange-rate series: first results for the Peseta-U.S. dollar case. Econ. Lett. **39**, 207–211 (1992)

Balanov, A., Janson, N., Postnov, D., Sosnovtseva, O.: Synchronization: From Simple to Complex. Springer, Heildelberg (2009)

Breakspear, M., Terry, J.R.: Topographic organization of nonlinear interdependence in multichannel human EEG. NeuroImage **16**, 822–835 (2002)
Brock, W.A., Dechert, W.: A test for independence based on the correlation dimension. Econom. Rev. **15**, 197–235 (1996)
Broomhead, D.S., King, Gregory P.: Extracting qualitative dynamics from experimental data. Physica D **20**, 217–236 (1986)
Cao, L.: Practical method for determining the minimum embedding dimension of a scalar time series. Physica D **110**, 43–50 (1997)
Cao, L., Mees, A.: Dynamics from multivariate time series. Physica D **121**, 75–88 (1998)
Cao, L., Hong, Y., Zhao, H., Deng, S.: Predicting economic time series using a nonlinear deterministic technique. Comput. Econ. **9**, 149–178 (1996)
Cao, L., Kim, B.G., Kurths, J., Kim, S.: Detecting determinism in human posture control data. Int. J. Bifurcat. Chaos **8**, 179–188 (1998a)
Cao, L., Mees, A., Judd, K., Froyland, G.: Determining the minimum embedding dimensions of input-output time series data. Int. J. Bifurcat. Chaos **8**, 1491–1504 (1998b)
Cao, L., Soofi, A.: Nonlinear deterministic forecasting of daily dollar exchange rates. Int. J. Forecast. **15**, 421–430 (1999)
Das, A., Das, P.: Chaotic analysis of the foreign exchange rates. Appl. Math. Comput. **185**, 388–396 (2007)
Devaney, R.L.: An Introduction to Nonlinear Dynamical Systems, 2nd edn. Addison-Wesley, Redwood City (1989)
Efron, B.: The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, Philadelphia (1982)
Fraser, A., Swinney, H.: Independent coordinates for strange attractors from mutual information. Phys. Rev. A **33**, 1134–1140 (1986)
Fujisaka, H., Yamada, T.: Stability theory of synchoronised motion in coupled-oscillator system. Prog. Theor. Phys. **69**, 32–47 (1983)
Galka, A.: Topics in Nonlinear Time Series Analysis: With Implications for EEG Analysis. World Scientific Publishing Company, Singapore (2000)
Galka, A., Ozaki, T.: Testing for nonlinearity in high-dimensional time series from continuous dynamics. Physica D **158**, 32–44 (2001)
Grassberger, P., Procaccia, I.: Measuring the strangeness of the strange attractors. Physica D **9**, 189–208 (1983)
Grassberger, P., Hegger, R., Kanz, H., Schaffrath, C., Screiber, T.: On noise reduction methods for chaotic data. Chaos **3**, 127–141 (1993)
Harvey, D., Leybourne, S., Newbold, P.: Testing the equality of prediction mean squared errors. Int. J. Forecast. **13**, 281–291 (1997)
Henon, M.: A two-dimensional mapping with a strange attractor. Commun. Math. Phys. **50**, 69–77 (1976)
Hilborn, R.C.: Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers. Oxford University Press, Oxford (1994)
Hinich, M.J.: Testing for gaussianity and linearity of a stationary time series. J. Time Ser. Anal. **3**, 169–176 (1982)
Hsieh, D.A.: Chaos and nonlinear dynamics: application to financial markets. J. Financ. **46**, 1839–1877 (1991)
Kantz, H., Schreiber, T.: Nonlinear Time series analysis. Cambridge University Press, Cambridge (1997)
Kaplan, D.: Math for biomedical engineering: activities for nonlinear dynamics signal processing. http://www.macalester.edu/kaplan/ (2004)
Kennel, M., Brown, R., Abarbanel, H.: Determining embedding dimension for phase space reconstruction using a geometrical reconstruction. Phys. Rev. A **45**, 3403–3411 (1992)
Kohzadi, N., Boyd, M.S.: Testing for chaos and nonlinear dynamics in cattle prices. Can. J. Agr. Econ. **43**, 475–484 (1995)

Kugiumtzis, D.: Surrogate data test on time series. In: Soofi, A., Cao, L. (eds.) Modelling and Forecasting Financial Data: Techniques of Non- linear Dynamics. Kluwer Academic Publishers, Boston (2002)
Larsen, C., Lam, L.: Chaos and the foreign exchange market. In: Lam, L., Naroditsky, V. (eds.) Modeling Complex Phenomena. Springer, New York (1992)
Lisi, F., Medio, A.: Is a random walk the best exchange rate predictor? Int. J. Forecast. **13**, 255–267 (1997)
Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**, 130–141 (1963)
May, R.: Simple mathematical models with very complicated dynamics. Nature **261**(5560), 459–467 (1976)
McLeod, A.I., Li, W.K.: Diagnostic checking ARMA time series models using squared-residual autocorrelations. J. Time Ser. Anal. **4**, 269–73 (1983)
Osborne, A., Provencale, A.: Finite correlation dimension for stochastic systems with power-law spectra. Physica D **35**, 357–381 (1989)
Ott, E., Sauer, T., Yorke, J.A.: Coping with chaos: analysis of chaotic data and the exploitation of chaotic systems. Wiley, New York (1994)
Pecora, L., Carroll, T.: Synchronization in chaotic systems. Phys. Rev. Lett. **64**, 821–824 (1990)
Pesin, Y.: Characteristic exponents and smooth ergodic theory. Russ. Math. Surv. **32**, 55–114 (1977)
Rulkov, N.F., Sushchik, M.M., Tsimring, L.S., Abarbanel, H.D.I.: Generalized synchronization of chaos in directionally coupled chaotic systems. Phys. Rev. E **51**, 980–994 (1995)
Sauer, T., Yorke, J.A., Casdagli, M.: Embodology. J. Stat. Phys. **65**, 579–616 (1991)
Scheinkman, J.A., LeBaron, B.: Nonlinear dynamics and stock returns. J. Bus. **62**, 311–337 (1989)
Schiff, S., So, P., Chang, T., Burke, R., Sauer, T.: Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. Phys. Rev. E **54**, 6708–6724 (1996)
Schreiber, T.: Extremely simple nonlinear noise reduction method. Phys. Rev. E **47**, 2401–2404 (1993)
Schreiber, T.: Interdisciplinary application of nonlinear time series methods. Phys. Rep. **320**, 1–86 (1998)
Screiber, T.: Constrained randomization of time series data. Phys. Rev. Lett. **80**, 2105 (1999)
Schreiber, T., Schmitz, A.: Improved surrogate data for nonlinearity tests. Phys. Rev. Lett. **77**, 635–638 (1996)
Soofi, A., Cao, L.: Nonlinear deterministic forecasting of daily peseta-dollar exchange rate. Econ. Lett. **62**, 175–180 (1999)
Soofi, A., Cao, L. (eds.): Modelling and forecasting financial data: techniques of nonlinear dynamics. Kluwer Academic Publisher, Boston (2002a)
Soofi, A., Cao, L.: Nonlinear forecasting of noisy financial data. In: Soofi, A., Cao, L. (eds.) Modelling and Forecasting Financial Data: Techniques of Nonlinear Dynamics. Kluwer Academic Publisher, Boston (2002b)
Soofi, A., Galka, A.: Measuring the complexity of currency markets by fractal dimensional analysis. Int. J. Theor. Appl. Financ. **6**, 553–563 (2003)
Soofi, A., Li, Z., Hui, X.: Nonlinear interdependence of Chinese stock markets. Quant. Financ. **12**, 397–410 (2012)
Strogatz, S.: Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering. Perseus Books, Cambridge (1994)
Takens, F.: Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence. Lecture notes in mathematics, vol. 898, pp. 366–381. Springer, Berlin (1981)
Theiler, J.: Spurious diemnsions from correlation algorithms applied to limited time series data. Phys. Rev. A **34**, 2427–2432 (1986)
Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D.: Testing for nonlinearity in time series: the method of surrogate data. Physica D **58**, 77–94 (1992)
Tsay, R.S.: Nonlinearity tests for time series. Biometrika **73**, 461–466 (1986)

Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. Physica D **35**, 395–424 (1989)
Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A.: Determining Lyapunov exponents from a time series. Physica D **16**, 285–317 (1985)
Yang, S., Brorsen, B.W.: Nonlinear dynamics of daily futures prices: conditional heteroscedasticity or chaos? J. Futures Markets **13**, 175–191 (1993)
Zhang, Y., Soofi, A., Wang, S.: Testing for nonlinearity of exchange rates: an information theoretic approach. J. Econ. Stud. **38**, 637–657 (2011)
Zhu, Q., Lu, L., Wang, S., Soofi, A.: Causal linkages among Shanghai, Shenzhen, and Hong Kong stock markets. Int. J. Theor. Appl. Financ. **7**, 135–149 (2003)