Are Financial Markets Stochastic: A Test for Noisy Chaos

Ahmed BenSaïda
University of Monastir, Faculty of Economics and Management of Mahdia
Sidi Massouad, Hiboun 5011, Mahdia, TUNISIA.

Abstract

As opposed to stochastic dynamics, recent studies suggested that financial markets might be governed by chaotic dynamics. Models that tried to explain market behavior are based on the stochastic hypothesis, which is observed when adding the perturbation error. However, stochastic models provide poor forecasts of the market, so far, which raises the question about the validity of the stochastic hypothesis. This paper presents a practical framework to test chaotic dynamics even for noisy systems as opposed to stochastic dynamics. It elaborates an easy-to-use and comprehensive algorithm to build a program to test chaos based on theoretical studies. Monte-Carlo simulations have confirmed that this test is powerful in detecting chaotic dynamics. The applications have confirmed that stock index returns S&P 500, Nikkei 225 and CAC 40 are stochastic and not chaotic.

Keywords: Chaos, Lyapunov, Nonlinear Dynamics.

JEL classification: C12, C13, C15, C45, C87.

Introduction

Chaos is a recent field of study which has been observed in physics since 1960s. The predominant linear techniques at that time were unable to explain some specific phenomena such as the movement of a driven pendulum where it may behave erratically and show irregular sequences of left and right turns, streams in the ocean, and more recently in meteorological science. As opposed to linear dynamics, these phenomena are called nonlinear dynamics. Recently, in finance, the debate still stands trying to find the answer whether stock movements are primary generated by stochastic or chaotic dynamics. These two systems look almost the same and even the powerful BDS test for IID cannot separate them. Consequently, two main streams describing the stock behavior have seen the light, and until then the distinction between them depends on the theoretical context. Moreover, recent tests to detect chaos are efficient only for correctly measured observation, and are not valid for financial data where data are sensitive to measurement noise. This paper presents a practical test for chaos which is valid even for noisy observation. Section 1 defines the chaos, section 2 defines the Lyapunov exponent, section 3 contains the procedure to estimate the Lyapunov exponent, section 4 gives an approximation to the chaotic map, section 5 describes the choice of parameters (L, m, q), section 6 gives the asymptotic distribution of the estimated Lyapunov exponent, section 7 contains simulations, section 8 investigates the dynamics of three major stock indexes, and finally we conclude.

1. Definition of chaos

In a scientific context, the word *chaos* has a slightly different meaning than it does in its general usage as *a state of confusion, lacking any order*. Chaos, with reference to *chaos theory,* refers to an apparent lack of order in a system that nevertheless obeys particular laws or rules; this understanding of chaos is synonymous with *dynamical instability*, a condition discovered by the physicist Henri Poincaré in the early 20^{th} century that refers to an inherent lack of predictability in some physical systems. The two main components of chaos theory are the ideas that systems - no matter how complex they may be - rely upon an underlying order, and that very simple or small systems and events can cause very complex behaviors or events. This latter idea is known as *sensitive dependence on initial conditions*.

Broadly speaking, mathematical models can be classified as either deterministic or stochastic models.

A deterministic process is a process that when repeated exactly in the same way will yield exactly the same outcome, in contrast to a stochastic process which yields different outcomes when repeated exactly in the same way. For example, if we go back in time and set *exactly* the same conditions as what we had observed in its very tiny detail, according to deterministic approach, the movement of the stock index would be exactly in the way as we observed it today, *citrus paribus*. Conversely, and according to the stochastic approach, even if we have exactly the same initial conditions as we had observed in the past, the stock index movement would generate a different pattern. One can suggest that deterministic systems exhibit only regular behavior, and if we are able to know the system's current state, we can easily predict its exact future state. However, this is not the case every time. Indeed, many deterministic systems exhibit irregular, random-like and unpredictable behavior known as the *butterfly effect*.

Chaos is defined in a time evolution described by a dynamical system by the solution of the following Ordinary Differential Equation ODE in continuous time: $\frac{\partial X_t}{\partial t} = F\left(X_t\right)$, or ODE in discrete time: $X_t = F\left(X_{t-1}\right)$, where $X_t \in \mathbb{R}^d$ and F is a $\mathbb{R}^d \to \mathbb{R}^d$ map. For example, the logistic map $x_t = \alpha \cdot x_{t-1}\left(1 - x_{t-1}\right)$ is known to exhibit chaotic behavior for $3.57 \leq \alpha \leq 4$ (Devaney, 1989). A higher dimensional example is given by the Hénon map: $F\left(x_t, y_t\right) = \left(1 - a \cdot x_{t-1}^2 + y_{t-1},\ b \cdot x_{t-1}\right)$. Numerical study shows that the Hénon map has complicated dynamics for $a = 1.4$ and $b = 0.3$.

Chaotic systems need sophisticated computational means (computers) to be correctly studied; as a result, the amazingly irregular behavior of some nonlinear deterministic systems was not appreciated. And when such systems are encountered in observations, they were explained as stochastic.

There are many studies of the mathematical aspects of chaos and dynamical systems, including Eckmann & Ruelle (1985) and Devaney (1989). Numerical implementations are discussed in Parker & Chua (1989). Chaos has attracted attention of the statistical community; see the special issue of the Journal of Royal Statistical Society series B, which includes Casdagli (1992), Smith (1992), and Nychka et al. (1997). Recognizing and quantifying chaos in time series was the subject of many studies. In fact, several approaches have been proposed including estimating fractal dimensions, Smith (1992), nonlinear forecasting, Casdagly (1992), estimating entropy (defined as the average rate that information is produced), Eckmann & Ruelle (1985), and estimating Lyapunov exponents, Wolf et al. (1985), Abarbanel et al. (1991) and Nychka et al. (1997).

Among the methods proposed, fractal dimension estimation is perhaps the simplest one. It provides a test about the finite dimensionality of a system. However, the dimension estimates is highly sensitive to measurement error in the data and may get worse with dynamical noise (Smith, 1992). Similar difficulty exists in the entropy estimates (Eckmann & Ruelle, 1985). Nonlinear forecasting is a more general concept because it includes nonlinearity in both deterministic and stochastic systems and it can be detected by the BDS test (Brock et al., 1996). Yet, the purpose of chaos test is to make difference between chaotic behavior and stochastic behavior. The problem encountered in fractal dimension estimation and entropy estimates is avoided in the Lyapunov exponent approach.

2. *The Lyapunov exponent*

Consider two points in a state space: X_0 and $X_0 + \Delta x_0$, each of them will generate an orbit in that space using some equation or system of equations. These orbits can be thought as parametric functions of a variable which is related to time. If we use one of the orbits as reference orbit, then the separation between the two orbits will also be a function of time. Because sensitive dependence can arise only in some portions of a system (like the logistic equation), this separation is also a function of the location of the initial value and has the form $\Delta x(X_0, t)$. In a system with attracting fixed points or attracting periodic points, $\Delta x(X_0, t)$ decreases asymptotically with time. If a system is unstable, then the orbits diverge exponentially for a while, but eventually settle down. For chaotic points, the function $\Delta x(X_0, t)$ will behave erratically. The perturbation Δx_0 created initially between X_0 and $X_0 + \Delta x_0$, generates perturbed and unperturbed trajectories, the difference between the two trajectories after t time steps is measured by:

$$\lambda = \lim_{t \to \infty} \frac{1}{t} \ln \frac{\left| \Delta x \left(X_0, t \right) \right|}{\left| \Delta x_0 \right|} \quad \text{(Eq. 1)}$$

This number, called the Lyapunov exponent λ, is useful for distinguishing among the various types of orbits. It works for discrete as well as continuous systems. It measures the average exponential divergence (positive exponent) or convergence (negative exponent) rate between nearby trajectories within a time horizon that differ in initial conditions only by an infinitesimally small amount. We distinguish 3 cases of λ:

- $\lambda < 0$: the orbit attracts to a stable fixed point or stable periodic orbit. Negative Lyapunov exponents are characteristic of dissipative or non-conservative systems. Such systems exhibit asymptotic stability; the more negative the exponent, the greater the stability. Super-stable fixed points and super-stable periodic points have a Lyapunov exponent of $\lambda = -\infty$.
- $\lambda = 0$: the orbit is a neutral fixed point (or an eventually fixed point). A Lyapunov exponent of zero indicates that the system is in some sort of steady state mode. A system with this exponent is conservative. Such systems exhibit Lyapunov stability. A system with a zero Lyapunov exponent is near the "*transition to chaos*" (Ellner & Turchin, 1995).
- $\lambda > 0$: the orbit is unstable and chaotic. Nearby points, no matter how close, will diverge to any arbitrary separation. All neighborhoods in the phase space will eventually be visited. These points are said to be unstable.

For an *n*-dimensional mapping, Oseledec's (1968) theorem states that the Lyapunov exponents are given by:

$$\lambda_i = \lim_{N \to \infty} \ln \left| v_i \left(N \right) \right| \quad \text{(Eq. 2)}$$

v_i's are the eigenvalues of the map matrix or the Jacobian product of F.

3. Estimating the Lyapunov exponent

Estimating the Lyapunov exponent is far from straightforward. While the theoretical concept provides a strong proof of its power in detecting chaotic dynamics ($\lambda \geq 0$), application to a time series is not evident. Wolf et al. (1985) have developed the first practical test of chaos on experimental data. Their algorithm is based on arbitrarily defining the Ordinary Differential Equations of the chaotic system, and test whether the data process coincides with the predefined system or not. This direct approach has limited applications since it bounds the chaotic behavior to the tested ones. Moreover, it cannot accept measurement errors or noise.

Perfectly chaotic systems are hard to model, first because of measurement errors, and second because the chaotic map is usually unknown and its approximation yields some perturbations (Schreiber & Kantz, 1995). Thus, a pure chaotic system is almost rare in reality, and we usually observe noisy systems. Given a set of time series $\left\{ x_t \right\}_{t=1}^{T}$, a noisy chaotic system can be written as:

$$x_t = f\left(x_{t-L}, x_{t-2L}, \cdots, x_{t-mL} \right) + \varepsilon_t \quad \text{(Eq. 3)}$$

The state-space representation of this equation is:

$$F : \begin{bmatrix} x_{t-L} \\ x_{t-2L} \\ \vdots \\ x_{t-mL} \end{bmatrix} \to \begin{bmatrix} x_t = f\left(x_{t-L}, x_{t-2L}, \cdots, x_{t-mL} \right) + \varepsilon_t \\ x_{t-L} \\ \vdots \\ x_{t-mL+L} \end{bmatrix} \quad \text{(Eq. 4)}$$

Where: ε_t represents the added noise. This formulation is a general form of the chaotic map, the amplitude of the added noise is measured by the variance of ε_t.

A noise-free system has $Var\left(\varepsilon_t\right)=\sigma_\varepsilon^2=0$, as $Var(\varepsilon_t)$ increases, the noise increases. In order to detect chaos, the amplitude of the noise should not exceed a certain limit $\sigma_{\lim}^2$, which is a function of the chaotic map.[1] If the noise exceeds that limit, the chaotic dynamics will be immersed in the noise, and it will be impossible to detect chaos, the system becomes then stochastic. The Lyapunov exponent for a noisy system λ_σ with $\sigma_\varepsilon^2>0$ will tend to the Lyapunov exponent of the deterministic skeleton λ_0 with $\sigma_\varepsilon^2=0$. It is suspected that $\lim_{\sigma_\varepsilon^2\to 0}\lambda_\sigma=\lambda_0$ and $\lambda_\sigma\le\lambda_0$. *L* is the *time delay* and its introduction allows the possibility of skipping samples during the reconstruction. The parameter *m* is the *embedding dimension* or the length of past dependence. Since the dynamics are unknown, we cannot reconstruct the original map that gave rise to the time series. Instead, we seek an embedding space where we can reconstruct the map from the observed data that preserve the invariant characteristics of the original unknown map. The embedding dimension *m* is in general different from the unknown dimension *d* (defined above). The simplest method for deriving a state vector $\left\{x_t\right\}_{t=1}^T$ is the *delay coordinates* proposed by Packard et al. (1980), *i.e.*, by using *L* and *m* instead. Takens' (1981) theorem states that reconstructed state vector will have the same dynamical properties as the original system if *m* is large enough ($m > 2d$).

Two main methods exist for the estimation of the largest Lyapunov exponent, the first one was proposed by Wolf et al. (1985) which is based on a direct approach. And the second one proposed by Eckmann & Ruelle (1985) which is based on a Jacobian approach. The direct approach consists of tracing the exponential divergence of nearby trajectories. In the presence of noise, however, the deterministic divergence is concealed by the noise process on the small scales. The direct approach is subject to many critics, because it requires long data series and is sensitive to dynamic noise (McCaffrey et al., 1992, Schreiber & Kantz, 1995, and Nychka et al., 1997).

The Jacobian-based approach can give consistent estimates of the Lyapunov exponents even in the presence of noise (McCaffrey et al., 1992, and Nychka et al., 1997). It consists of computing the Jacobian matrix of the chaotic map *F*:

$$J_t=\begin{pmatrix}\frac{\partial f}{\partial x_{t-L}} & \frac{\partial f}{\partial x_{t-2L}} & \cdots & \frac{\partial f}{\partial x_{t-mL+L}} & \frac{\partial f}{\partial x_{t-mL}}\\ 1 & 0 & \cdots & 0 & 0\\ 0 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 1 & 0\end{pmatrix} \quad \text{(Eq. 5)}$$

Because the Lyapunov exponents measure the log-difference in the norms of the two trajectories after *M* time steps, Eckmann & Ruelle (1985) have shown that for small perturbations, the linear approximation will be defined by the derivatives of the map *F* (relative to *X*):

$$\lambda=\lim_{M\to\infty}\frac{1}{M}\ln\left\|\prod_{t=1}^{M-1}J_{M-t}\right\| \quad \text{(Eq. 6)}$$

Where $\|\ \|$ indicates the Euclidian vector norm. The Euclidian norm is equivalent to the square root of the largest eigenvalue of the matrix transposed and multiplied by itself, *i.e.*, $\|X\|=\max\left(eig\left(X'X\right)^{1/2}\right)$. Posing $T_M=\prod_{t=1}^{M-1}J_{M-t}$, the dominant Lyapunov exponent could then be written as:

[1] A further research on how to determine $\sigma_{\lim}^2$ is suggested.

$$\lambda = \lim_{M \to \infty} \frac{1}{2M} \ln(v_1) \quad \text{(Eq. 7)}$$

Where v_1 is the largest eigenvalue of the matrix $T_M{}'T_M$: $v_1 = \max\left(eig\left(T_M{}'T_M\right)\right)$. Since the data are finite, an estimate of the Lyapunov exponent is: $\hat{\lambda} = \frac{1}{2M}\ln(v_1)$, where M is the number of evaluation points or the *block-length* ($M \leq T$). McCaffrey et al. (1992) suggest that the evaluation points should be equally spaced. Moreover, Shintani & Linton (2004) have found that the best magnitude for M was close to $T^{1/3}$ for processes with chaotic like behavior (M can also cover the full sample, *i.e.*, $M = T$).

The estimated $\hat{\lambda}$ tends to have a positive systematic bias compared to the theoretical Lyapunov exponent defined in (Eq.7), because one condition is that M tends to infinity, and for limited M, the obtained quantity is over-estimated. Hence, one possible correction is to compute the Orthogonal-triangular decomposition estimate of λ as proposed by Abarbanel et al. (1991), *i.e.*, by multiplying T_M by a unit vector U_0 to reduce the systematic positive bias in the formal estimate of λ. U_0 is chosen at random with respect to uniform measure on the unit sphere, and then the estimated λ converges asymptotically to the global Lyapunov exponent as $M \to \infty$, because U_0 has zero probability of falling into the subspace corresponding to subdominant exponents (Nychka et al., 1997, and Bailey et al., 1998). In practice, however, U_0 is chosen as: $U_0 = (1, 0, \ldots, 0)'$.

4. Approximating the chaotic map F

The procedure described above is based on the estimation of the Jacobian matrix of the chaotic map F. However, for a scalar time series $\{x_t\}_{t=1}^{T}$, the map generating the process is usually unknown; as a result, the Jacobian matrix could not be estimated and we cannot compute the Lyapunov exponent. For that purpose, we need to approximate the unknown chaotic map with a known function that can *learn* the process by *reading* the relation $y = G(x)$, where $x, y \in \mathbb{R}^p$. Some candidate functions exist, *e.g.*, splines (smooth piecewise polynomials), neural network, nearest neighbor …*etc*. Hornik et al. (1989) proposed that special networks can, in principle, approximate any smooth, nonlinear function to arbitrary accuracy as the number of hidden units goes to infinity.

Initially, neural networks were developed as a simulation model of the brain. A neural net system consists of neurons (cells), neural interconnection (internal links), and connections with the outer world. In a multi-layer network, neurons are organized in layers with interconnections only between cells of neighboring layers. The network is connected to the outer world by the first or input layer and by the last or output layer. The space existing between input and output is called *hidden layers*. An input signal is fed forward through the hidden layers toward the output layer without feedback. The number of cells in a layer is called the *dimension of the layer*. The basic concept of feed-forward neural net is the propagation of a signal from one layer to another.

The hidden layer learns to *recode* (or to *provide a representation* for) the inputs. More than one hidden layer can be used. The architecture of multiple hidden layers is more powerful than single-layer networks: it can be shown that any mapping can be learned, given two hidden layers.

The units are a little more complex than those in the original *perceptron* (a simple form of neural networks. They have no hidden layers, and can only perform linear classification tasks): their input/ output graph is represented by an activation function.

The neural net receives a number of inputs either from original data, or from the output of other neurons in the neural network. Each input comes via a connection that has a strength (or *weight*). Each neuron also has a single *threshold value*. The weighted sum of the inputs is formed, and the threshold subtracted to compose the *activation* of the neuron. The activation signal is passed through an *activation function* (also known as a *transfer function*) to produce the output of the neuron. A simple network has a feed-forward structure: signals flow from inputs, forwards through any hidden units, eventually reaching the output units. Such a structure has stable behavior.

However, if the network is *recurrent* (contains connections back from later to earlier neurons) it can be unstable, and has very complex dynamics. Recurrent networks are very interesting to researchers in neural networks, but so far it is the feed-forward structures that have proved most useful in solving real problems (Haykin, 1998).

A typical feed-forward network has neurons arranged in a distinct layered topology. The input layer is not really neural at all: these units simply serve to introduce the values of the input variables. The hidden and output layer neurons are each connected to all of the units in the preceding layer. Again, it is possible to define networks that are partially-connected to only some units in the preceding layer. However, for most applications, fully-connected networks are better. When the network is executed, the input variable values are placed in the input units, and then the hidden and output layer units are progressively executed. Each of them calculates its activation value by taking the weighted sum of the outputs of the units in the preceding layer, and subtracting the threshold. The activation value is passed through the activation function to produce the output of the neuron. When the entire network has been executed, the outputs of the output layer act as the output of the entire network.

Let $\left(x_{t-L}, x_{t-2L}, \cdots, x_{t-mL}\right)$ be the input signal of an *m*-dimensional layer, this signal is transmitted to a *q*-dimensional layer with connection weights $\beta_{i,j}$, $i = 1, \ldots, m$ and $j = 1, \ldots, q$, so the j^{th} cell will receive a total signal of $\sum_{i=1}^{m} \beta_{i,j} x_{t-iL}$, which is the total sum of signals provided by the foregoing *m* cells. Each cell has its own sensitivity called *internal threshold*, it is moreover incorporated in the form of an additional signal $\beta_{0,j}$. Thus, each cell will receive in total: $\beta_{0,j} + \sum_{i=1}^{m} \beta_{i,j} x_{t-iL}$. The propagation of a signal is governed by an activation function of a cell denoted: Ψ. Typically, Ψ is a sigmoid function which is monotone and bounded such that $\lim_{u\to\infty} \Psi(u) = 1$ and $\lim_{u\to-\infty} \Psi(u) = 0$. However, to obtain the full power of neural net approximation, these limits could not be respected (Hornik et al., 1994). Usually, the logistic function $\Psi(u) = \frac{1}{1+e^{-u}}$ is used. More powerful function, however, exits; mainly the hyperbolic tangent $\Psi(u) = \tanh(u)$. Its lower limit is -1 instead of zero.

The output signal from the j^{th} cell is then $\Psi\left(\beta_{0,j} + \sum_{i=1}^{m} \beta_{i,j} x_{t-iL}\right)$, the total *q* hidden layers of the neural net, with α_j inter-layers connection weights and α_0 the network threshold, will transmit $\alpha_0 + \sum_{j=1}^{q} \alpha_j \Psi\left(\beta_{0,j} + \sum_{i=1}^{m} \beta_{i,j} x_{t-iL}\right)$. The chaotic map *F* could then be approximated by:

$$x_t \approx \alpha_0 + \sum_{j=1}^{q} \alpha_j \Psi\left(\beta_{0,j} + \sum_{i=1}^{m} \beta_{i,j} x_{t-iL}\right) + \varepsilon_t \qquad \text{(Eq. 8)}$$

It is possible to add some other fitting functions to the above formulation, such as polynomial formulation or splines. Nevertheless, the added functions will increase the number of coefficients to be estimated, which will render any approximation impractical.

The noise $\{\varepsilon_t\}_{t=mL+1}^{T}$ should be minimized to compensate the loss of information engendered from the approximation by the neural network. Consequently, an obvious estimation algorithm is the nonlinear least square NLS which minimizes: $S(\theta) = \sum_{t=mL+1}^{T} \varepsilon_t^2 = \sum_{t=mL+1}^{T} \left[x_t - f\left(\underline{x_{t-L}}, \theta\right)\right]^2$, where θ represents the parameters to estimate.

5. Choice of the parameters L, m and q

The triplet (*L*, *m*, *q*) defines the complexity of the chaotic map. In theory, as the number of hidden layers *q* goes to infinity, neural net function can approximate any smooth, nonlinear function to arbitrary accuracy. A choice of the time delay *L* that keeps time dependence in the scalar time series is desirable. However, a too large value causes a loss of information, and conversely, a value too small makes the observations vector temporarily close and remote, giving rise to uncertainties.

For the embedding dimension *m*, it should satisfy Takens' (1981) theorem: $m > 2d$. Yet, one should know the dimension of the original state space *d* to determine the necessary dimension of the reconstructed state space *m*. One approach to find *m* is the *singular system approach* by Broomhead & King (1986)[2] widely used in many areas of applied numerical linear algebra. The method suggests an initial reconstitution of the state space with an arbitrarily large *m*, even larger than suggested by Takens' theorem.

The common handicap in choosing the triplet (*L*, *m*, *q*) is that low parameters may prevent the neural network from reasonably approximating the map that generates the scalar time series. On the other hand, large parameters increase computational time exponentially because the number of coefficients to estimate will increase. Nychka et al. (1997) have adopted a strategy to select the triplet (*L*, *m*, *q*) which minimizes the Schwarz (or Bayesian) Information Criterion as defined by Schwarz (1978), $SIC = -2L\left(\hat{\theta}\right) + n\ln\left(T\right)$, where $L(\hat{\theta})$ is the estimated log-likelihood function, *T* is the sample size, and *n* is the number of estimated coefficients. However, this strategy tends to eliminate higher-order regressions where *L*, *m* or *q* are rather high; hence, medium to high complex chaotic dynamics could not be revealed because the SIC penalizes models as the number of coefficients increases. Besides, this method eliminates the same chaotic map even when we change the starting value.

Hypothetically, increasing the order of the triplet (*L*, *m*, *q*) allows the neural net function to approximate the map to arbitrarily accuracy. If, for any given combination of the triplet (*L*, *m*, *q*), a positive Lyapunov exponent is obtained, this provides a strong indication of the presence of chaos. Conversely, the failure to obtain positive Lyapunov exponent for any combination of (*L*, *m*, *q*), rejects the hypothesis of chaotic dynamics.

One may wonder then whether choosing sufficiently high order parameters (*L*, *m*, *q*) is adequate enough to conduct the estimation, since higher complex neural net would cover lower complex functions. However, even higher orders will cover the optimal order's combination and add unwanted coefficients structure with increasing lagged values of the times series, which will result in losing information (from the time series), and facing the problem of redundant variables.

The practical procedure to estimate the Lyapunov exponent, is to choose a ceiling for the triplet (*L*, *m*, *q*) not to exceed, perform nonlinear least square estimation of the models for all parameters (*L*, *m*, *q*), which yields in total $L*m*q$ regressions, compute the Lyapunov exponent for each regression, and finally choose the triplet (*L*, *m*, *q*) for which the higher exponent λ is obtained. This method is computationally demanding, yet, it insures the consistency of the result. The main drawback of this method is that when the dynamics are generated by a high-level chaos, under the chosen ceiling triplet (*L*, *m*, *q*), the test will not reveal any chaotic dynamics because the tested chaotic maps lie beneath the ceiling triplet. In theory, increasing the ceiling triplet (*L*, *m*, *q*) to infinity will increase the power of the test; however, in practice the ceiling triplet should have reasonable values to enable computation with current computers.

6. Asymptotic distribution of the estimated Lyapunov exponent

The introduction of noise in the aforementioned chaotic map *F*, should in principle affect the accuracy of the estimated Lyapunov exponent by making it stochastic. Until recent years, the Lyapunov exponents were computed as a constant measure (Wolf et al., 1985). Some works on the asymptotic distribution of λ were carried out by Nychka et al. (1997), Bailey et al. (1998), and Shintani & Linton (2004).

[2] This method has many names in the literature including: *principal component analysis*, *factor analysis*, and *Karhunen-Loeve decompositions*.

Nychka et al. (1997) and Bailey et al. (1998) have used a central limit theorem from a functional Markov process to obtain distributional results for the local Lyapunov exponent process. Under some conditions, mainly that F is bounded, its Jacobian is also bounded, ε_t are IID and their probability distribution function is bounded; these conditions are needed to have a product of several Jacobian which yields a distribution that is not singular. In their theorem 4.1, Bailey et al. (1998) state that:

$$\frac{\sqrt{M}\left(\hat{\lambda}_M-\lambda\right)}{\sqrt{Var\left(\hat{\lambda}_M\right)}}\xrightarrow{Asymptotically}N\left(0,1\right)\text{ as }M\to\infty \qquad \text{(Eq. 9)}$$

Shintani & Linton (2004) have shown under some conditions that the variance of the i^{th} largest Lyapunov exponent is:

$$Var\left(\hat{\lambda}_i\right)=\Sigma_i=\lim_{M\to\infty}Var\left(\frac{1}{M}\sum_{t=1}^{M}\eta_{i,t}\right) \qquad \text{(Eq. 10)}$$

Where: $\eta_{i,t}=\omega_{i,t}-\lambda_i$, with $\omega_{i,t}=\frac{1}{2}\ln\left[\frac{v_i\left(T'_tT_t\right)}{v_i\left(T'_{t-1}T_{t-1}\right)}\right]$ for $t\geq 2$ and $\omega_{i,1}=\frac{1}{2}\ln\left[v_i\left(T'_1T_1\right)\right]$.

Since η_t are serially correlated and non IID, we will employ the Heteroskedasticity and Autocorrelation Consistent HAC covariance matrix estimator as described by Andrews (1991):

$$\hat{\Sigma}_i=\sum_{j=-M+1}^{M-1}\xi\left(\frac{j}{S_M}\right)\hat{\delta}\left(j\right)\text{ and }\hat{\delta}\left(j\right)=\frac{1}{M}\sum_{t=|j|+1}^{M}\hat{\eta}_t\hat{\eta}_{t-|j|} \qquad \text{(Eq. 11)}$$

Where: $\eta_{i,t}=\omega_{i,t}-\lambda_i$ with $\omega_{i,t}=\frac{1}{2}\ln\left[\frac{v_i\left(T'_tT_t\right)}{v_i\left(T'_{t-1}T_{t-1}\right)}\right]$ for $t\geq 2$ and $\omega_{i,1}=\frac{1}{2}\ln\left[v_i\left(T'_1T_1\right)\right]$.

$\xi(\ )$ and S_M denote a kernel function and a lag truncation parameter, respectively.

Assumption: (HAC estimation). $\xi:\mathbb{R}\to\left[-1,1\right]$ *is a piecewise continuous function, continuous and taking the value* 1 *at zero, symmetric around zero, and has a finite second moment. In other worlds, the class of kernel is*:

$$\Xi=\left\{\begin{array}{l}\xi(\ ):\xi\left(0\right)=1,\xi\left(-x\right)=\xi\left(x\right)\forall x\in\mathbb{R},\int_{-\infty}^{+\infty}\xi^2\left(x\right)dx<\infty,\\ \xi(\ )\text{ is continuous at 0 and all but a finite number of points}\end{array}\right\}$$

The lag truncation parameter must satisfy $\lim_{M\to\infty}S_M=\infty$ and $\lim_{M\to\infty}\frac{S_M}{M}=0$.

In a comparison between kernels based on Monte-Carlo simulation, Andrews (1991) suggested that the Quadratic Spectral “QS” kernel is the optimal choice among others:

$$\xi_{QS}\left(x\right)=\frac{25}{12\pi^2x^2}\left(\frac{\sin\left(\frac{6\pi x}{5}\right)}{\frac{6\pi x}{5}}-\cos\left(\frac{6\pi x}{5}\right)\right) \qquad \text{(Eq. 12)}$$

The optimal lag truncation parameter S_M for the QS kernel is: $S_M^*=1.3221\left(\kappa\cdot T\right)^{1/5}$. Andrews (1991) has discussed the procedure to estimate κ, which is rather difficult because it depends on the estimated function, for simplicity's sake I will choose κ = 1 since this approximation will not disagree in any way with the lag truncation parameter conditions stated by Andrews (1991).

The estimated variance of the Lyapunov exponent will converge in probability to the true variance under the aforementioned conditions.

The test hypothesis could then be constructed. The null hypothesis to test is H_0: $\lambda \geq 0$ against the alternative H_1: $\lambda < 0$. The rejection of the null hypothesis provides a strong evidence of no chaotic dynamics. The computed λ stands for the dominant Lyapunov exponent, which corresponds to the largest eigenvalue v_1. The test statistic to compute is:

$$\hat{W} = \frac{\hat{\lambda}_M}{\sqrt{\hat{\Sigma}_1 / M}} \rightarrow N(0,1) \qquad \text{(Eq. 13)}$$

The null hypothesis is rejected if $\hat{W} \leq -z_\alpha$, where z_α is the critical value that satisfies $\Pr[Z \geq z_\alpha] = \alpha$ with Z being a standard normal variable.

The magnitude of λ will depend on the degree of divergence provided by the chaotic map F, *i.e.*, as the map gives more and more divergent outputs, λ is expected to increase toward infinity. Similarly, very low chaotic dynamics (with lower divergent outputs) will display λ almost zero. A system with a zero Lyapunov exponent λ is near the "transition to chaos" (Ellner & Turchin, 1995). On the other hand, negative values of λ indicate the absence of chaotic dynamics in the map function. Purely random numbers are expected, theoretically, to have λ which tends toward minus infinity.

7. Simulation

To confirm the methodology in previous sections, some simulations should be carried on. Consequently, two chaotic models and three non-chaotic models are chosen. Noise is then added to the models. The chaotic models are the well-known tent-map and the logistic map; non-chaotic models are composed of a GARCH model, an ARMA model and a simple random number. The core part of the chaos test program written on MATLAB is provided in http://www.mathworks.com/matlabcentral/fileexchange/22667.

Model 1: The tent map: $\begin{cases} x_t = 2x_{t-1} + \varepsilon_t & \text{, if } x_{t-1} < 0.5 \\ x_t = 2(1 - x_{t-1}) + \varepsilon_t & \text{, if } x_{t-1} \geq 0.5 \end{cases}$ is the simplest chaotic map, with the starting value $x_0 = 0.7$ and $\varepsilon_t \rightarrow NID(0, \sigma_\varepsilon^2)$ a Normally and Independently Distributed white noise, with $\sigma_\varepsilon = 0.01$.

Model 2: Logistic map: $x_t = 3.57 x_{t-1}(1 - x_{t-1}) + \varepsilon_t$, with the starting value $x_0 = 0.7$ and $\varepsilon_t \rightarrow NID(0, \sigma_\varepsilon^2)$ a normally distributed white noise, with $\sigma_\varepsilon = 0.01$.

Model 3: GARCH(1, 1) process: $x_t = \varepsilon_t \sqrt{h_t}$, where $h_t = 0.01 + 0.1 x_{t-1}^2 + 0.85 h_{t-1}$ and $\varepsilon_t \rightarrow NID(0, \sigma_\varepsilon^2)$ Normally and Independently Distributed, with $x_0 = 0$ and $h_0 = \frac{0.01}{1 - (0.1 + 0.85)} = 0.2$ is the sample unconditional variance.

Model 4: ARMA(1, 1) process: $x_t = 0.1 + 0.2 x_{t-1} + \varepsilon_t + 0.15 \varepsilon_{t-1}$, with $x_0 = 0.1$ and $\varepsilon_0 = 0$. $\varepsilon_t \rightarrow NID(0, \sigma_\varepsilon^2)$ is a normally distributed white noise.

Model 5: Normally distributed random number with zero mean and unit variance: $x_t \rightarrow NID(0,1)$.

Some care should be taken in choosing random number generator *RNG*. The *RNG* used in this simulation combines a multiplicative integer congruential generator invented by Lehmer (1951) and an integer-shift-register generator to create uniformly distributed random numbers. Marsaglia's "*ziggurat method*" (Marsaglia & Tsang, 2000) is then applied to obtain normally distributed random numbers from the uniform random numbers.[3]

[3] Mathematically, ziggurats are two-dimensional step functions. A one-dimensional ziggurat underlies Marsaglia's algorithm.

This is done by using a simple uniform *RNG* to sample a table of pre-computed values which partition the normal distribution into regions of 1/32nd of its area. By comparison with the conventional method of obtaining random numbers by multiplicative congruential generation alone,[4] the combination of three different *RNG* methods employed here should make the resulting samples look almost perfectly independent.[5] The initial seed value of the *RNG* is set by the clock of the computer at the time the program was run.

For all models, 1000 observations are simulated. The activation function used in the neural network is the hyperbolic tangent, and the maximum parameters (L, m, q) values are (5, 6, 5). Simulation results are presented in Table 1.

Simulation results have confirmed the test by accepting H_0 of chaotic dynamics for chaotic models 1 and 2, and by rejecting the null hypothesis for other non-chaotic models 3, 4 and 5.

I changed the starting values for all models[6] (for model 5, I changed the mean and variance) to check the consistency of the results, the Lyapunov exponent λ changed slightly but the accepted hypotheses have remained unchanged (the model orders have also changed).

Concerning the chaotic models 1 and 2, increasing the variance of the added noise σ_ε^2 beyond a certain limit rejects the null hypothesis H_0 of chaos dynamics, and the dominant Lyapunov exponent becomes negative.[7] In this case, the stochastic behavior of the noise enfolds the deterministic behavior of the chaotic map, and the dynamics are converted to stochastic.

The Lyapunov exponent for the random numbers (model 5) is expected to tend toward minus infinity. However, although it is the smallest value among other results, it is still a finite number. What does this mean? Does the above described procedure concerning the test for chaos wrong? To answer this question, we must have a closer look at the generated "random numbers". Numbers generated by the computer are called *pseudorandom* numbers, because computers are in principle deterministic machines and should not exhibit random behavior. If the computer does not access some external device, like a gamma ray counter or a clock, then it must really be computing pseudorandom numbers. One favorite definition was given by Lehmer (1951):

"*A random sequence is a vague notion … in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests traditional with statisticians …*"

Pseudorandom numbers generated by all computers are based on the *RNG*, which is a specific algorithm to be executed by the program. The program needs a starting value for any execution, even in case of random numbers generation. The starting value is given by a state of the *RNG* which is, in the aforementioned method, a 35 length row-vector. Each time a pseudorandom number is generated, the state of the *RNG* changes accordingly and in a pre-specified way. The next pseudorandom number is generated accordingly to the new state, which in turn was formed from the previous pseudorandom number. The art of computer language makes that the generated pseudorandom numbers are independent; and the obtained pseudorandom numbers are almost purely random. The BDS test for independence is applied to the generated random numbers, and it has accepted the null hypothesis of IID-ness. Hence, the chaos test result for model 5, where data are composed of computer-generated random numbers, is acceptable.[8]

---

[4] Traditionally, normal random numbers are obtained by simple scaling of the uniform random numbers.

[5] The period of the above-mentioned uniform *RNG* is approximately 2^{1492}, *i.e.*, even if one drew random numbers at a rate of 20 million per second, non-stop, (possible on a Dual Core 2, 2.26 GHz PC), they would repeat themselves only after 2^{435} years!

[6] For the tent map and logistic map, starting value x_0 must lie strictly between 0 and 1: $0 < x_0 < 1$.

[7] I suggest further researches on how to determine the limiting variance of the stochastic noise $\sigma_{\lim}^2$.

[8] The pseudorandom numbers are generated using a special computer code, a starting value is necessary to jump-start the procedure. In theory, pseudorandom numbers form a highly complex chaotic map. So complex that any prediction becomes impossible, it can be then assimilated to as a stochastic dynamic.

Next simulation is conducted on models 1 and 2 with conditionally heteroskedastic noise:

Model 1#: The tent map: $\begin{cases} x_t = 2x_{t-1} + \varepsilon_t & \text{, if } x_{t-1} < 0.5 \\ x_t = 2(1 - x_{t-1}) + \varepsilon_t & \text{, if } x_{t-1} \geq 0.5 \end{cases}$, with the starting value x_0 = 0.7 and $\begin{cases} \varepsilon_t = u_t \sqrt{h_t} \;\; ; u_t \sim NID(0,1) \\ h_t = 1E-6 + 0.1\varepsilon_{t-1}^2 + 0.85h_{t-1} \end{cases}$ a GARCH(1, 1) noise, with ε_0 = 0 and $h_0 = \dfrac{1E-6}{1-(0.1+0.85)} = 2E-5$ is the unconditional sample variance (the unconditional standard deviation is $\sigma_\varepsilon = \sqrt{h_0} = 0.0045$).

Model 2#: Logistic map: $x_t = 3.57x_{t-1}(1 - x_{t-1}) + \varepsilon_t$, with the starting value x_0 = 0.7 and $\begin{cases} \varepsilon_t = u_t \sqrt{h_t} \;\; ; u_t \sim NID(0,1) \\ h_t = 5E-6 + 0.1\varepsilon_{t-1}^2 + 0.85h_{t-1} \end{cases}$ a GARCH(1, 1) noise, with ε_0 = 0 and $h_0 = \dfrac{5E-6}{1-(0.1+0.85)} = 1E-4$ is the unconditional sample variance (the unconditional standard deviation is $\sigma_\varepsilon = \sqrt{h_0} = 0.01$).

The unconditional variance of the added noise is chosen small. If the unconditional variance exceeds a certain limit, the stochastic behavior of the added noise will enfold the chaotic map, and the test will not detect chaotic dynamics. Results are in Table 2.

The test detects chaos even in presence of conditionally heteroskedastic noise. Hence, the methodology described in this chapter can distinguish between stochastic and chaotic dynamics even in presence of moderate noise. The conducted simulations have supported the theoretical concept of the test.

8. Application

The data is composed of three major stock indexes rate, which are the S&P 500, the Nikkei 225 and the CAC 40 from January 1st, 1999 until December 31st, 2008, which makes 9 years of daily observation. The data are downloaded from Yahoo! Finance. The rates are the daily adjusted closing spot indexes, which are transformed to obtain the returns.

Stock index returns are largely considered in the literature as stochastic, hence governed by some probability distribution, and the models applied to investigate the dynamics of the returns are based on the hypothesis of stochastic dynamics. The stochastic behavior of the index returns is present in the error term introduced in the models. GARCH class models are so far considered by the literature as the best to model indexes returns behavior; however, the poor forecasting power of these models has awaken the question about the true dynamics of market returns: stochastic or chaotic. Results are summarized in Table 3. The test has confirmed the hypothesis that stock indexes returns are stochastic and not chaotic.

Conclusion

The methodology described in this paper can distinguish between stochastic and chaotic dynamics even in presence of moderate noise. The conducted simulations have supported the theoretical concept of the test. Furthermore, the test has proved that stock indexes returns are stochastic and not chaotic.

Table 1: Chaos test simulation results

	(L, m, q)	λ	p-value*	CI**	Hypothesis
Model 1	(1, 2, 1)	0.1473	1	[0.1443, +∞[	H_0
Model 2	(1, 1, 1)	0.9150	1	[0.9098, +∞[	H_0
Model 3	(2, 6, 4)	-0.4638	5.1656E-90	[-0.502, +∞[	H_1
Model 4	(5, 5, 3)	-0.5306	3.167E-282	[-0.5549, +∞[	H_1
Model 5	(5, 4, 5)	-0.6324	4.8425E-65	[-0.6939, +∞[	H_1

* At 5% significance level, H_0 is rejected for p-values less than 0.05.
** Confidence Interval at 5% significance level.

Table 2: Chaos test results for noisy chaotic models

	(*L*, *m*, *q*)	λ	*p*-value*	CI**	Hypothesis
Model 1#	(1, 2, 1)	-0.0081	0.3345	[-0.0397, +∞[	H_0
Model 2#	(1, 1, 1)	0.9069	1	[0.9023, +∞[	H_0

* At 5% significance level, H_0 is rejected for *p*-values less than 0.05.
** Confidence Interval at 5% significance level.

Table 3: Chaos test results for the indexes returns

	(*L*, *m*, *q*)	λ	*p*-value*	CI**	Hypothesis
S&P 500	(2, 6, 1)	-0.3487	0	[-0.3607, +∞[	H_1
Nikkei 225	(4, 6, 4)	-0.4007	0	[-0.4182, +∞[	H_1
CAC 40	(5, 6, 4)	-0.3860	0	[-0.3999, +∞[	H_1

* At 5% significance level, H_0 is rejected for *p*-values less than 0.05.
** Confidence Interval at 5% significance level.

References

Abarbanel, H. D. I., Brown R., and Kennel M. B. (1991). Lyapunov exponents in chaotic systems: their importance and their evaluation using observed data. *International Journal of Modern Physics B*, 5(9), 1347-1375.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3), 817-858.

Bailey, B., Nychka D., and Ellner S. (1998). A central limit theorem for Local Lyapunov Exponents. Working paper.

Brock W., Dechert, D., Scheinkman, J., and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Review*, 15(3), 197-235.

Broomhead, D. S., and King J. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D*, 20(2), 217-236.

Casdagly, M. (1992). Chaos and deterministic versus stochastic nonlinear modeling. *Journal of Royal Statistical Society*, 54(2), 303-328.

Devaney, R. L. (1989). *An introduction to chaotic dynamical systems*. (2nd ed.). Cambridge Addison Wesley publishing company.

Eckmann, J., and Ruelle D. (1985). Ergodic theory of chaos and strange attractors. *Revue of Modern Physics*, 57(3), 617-656.

Ellner, S., and Turchin, P. (1995). Chaos in a noisy world: new methods and evidence from time series analysis. *American Naturalist*, 145(3), 343-375.

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. (2nd ed.). New Jersey: Prentice Hall.

Hornik, K., Stinchcombe M., and White H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.

Hornik, K., Stinchcombe M., White H., and Auer P. (1994). Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives. *Neural Computation*, 6(6), 1262-1275.

Lehmer, D. H. (1951). Mathematical methods in large-scale computing units. *Annals of the Computation Laboratory of Harvard University*, 26, 141-146.

Marsaglia, G., and Tsang W. (2000). The ziggurat method for generating random variables. *Journal of statistical software*, 5(i08), 1-7.

McCaffrey, D., Nychka D., Ellner S., and Gallant R. (1992). Estimating Lyapunov exponents with nonparametric regression. *Journal of American Statistical Society*, 87(419), 682-695.

Nychka, D., Ellner S., and Bailey B. (1997). Chaos with confidence: asymptotics and application of local Lyapunov exponents. *American Mathematical Society*, 115-133.

Oseledec, V. (1968). A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Moscow Mathematical Society*, 19, 197-231.

Packard, N. H., Crutchfield J. P., Farmer J. D., and Shaw R. S. (1980). Geometry from a time series. *Physical Review Letters*, 45(9), 712-716.

Parker, T. S., and Chua L. O. (1989). *Practical numerical algorithms for chaotic systems*. New York: Springer-Verlag.

Schreiber, T., and Kantz H. (1995). Noise in chaotic data: diagnosis and treatment. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1), 133-142.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.

Shintani, M., and Linton O. (2004). Nonparametric neural network estimation of Lyapunov exponents and direct test for chaos. *Journal of Econometrics*, 120(1), 1-33.

Smith, L. A. (1992). Identification and prediction of low dimensional dynamics. *Physica D*, 58(1-4), 50-76.

Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898, 366-381.

Wolf, A., Swift J. B., Swinney H. L., and Vastano J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D*, 16(3), 285-317.