How to distinguish Reality from Simulation ?

K. Bertels, L. Neuberg *

1 Introduction

When using simulations to study real world phenomena, we must make sure that the artificially generated data or model accurately reflects reality. Even though prediction is especially in the beta sciences the ultimate measure to evaluate the fitness of a model, in economics this has never been the case. However, the least one can expect is that a model allows to describe the overall behavior of the system and that similar properties are observed. A lot of emphasis has lately been put on non linear economic phenomena having sometimes also chaotic properties. When using simulation to study such phenomena, we must be sure that the computer based simulation possesses similar characteristics as its real world counterpart. In the case of chaotic systems, detecting the presence of chaos is not an easy and straightforward task and involves the use of computationally intensive methods. In this paper, we present a number of such techniques. Without going into the mathematical details of each of them, we illustrate their use, using both real world data as well as simulation based data.

The paper is organized as follows. In a first section, we first briefly describe the simulation model that was used to generate the 'artificial' time series. Consequently, we give a definition of chaos and then introduce the most currently used descriptive and numerical tools such as phase space, embedding dimension, correlation dimension and spectral analysis.

2 A Simulated Small Stock Exchange Market (SSSEM)

Before presenting the different tools, we first briefly describe the simulation model, developed using the Structured Data Modelling Language, which will be used to study a number of financial phenomena. The model we have constructed is very simple. It consists of only 5 agents trading one stock during a period of time. They can use money to clear the orders and they have a negociation mechanism in order to establish a price at which either party wants to sell or buy. The evolution of the intraday price of this stock is then the time series used in the numerical and descriptive tools.

3 Definition of Chaos

Only dynamical systems can exhibit chaotic behavior. In [BDVT91], a dynamical system is defined as a triple (M,T,ϕ), consisting of a phase space, a time set T and an evolution operator $\phi : M$ x $T \rightarrow M$ where the phase space **M** represents all the possible states of a system at any particular moment, the time set **T** is the time lag in which the system is defined and the evolution operator ϕ specifies deterministically from any given state the future of the system.

Furthermore, chaotic phenomena can appear in dynamic systems which have in addition the following characteristics [FE80] :

- non-linear : only non-linear equations exhibit the kind of dynamic behavior where phenomena such as chaos arise. In the linear system, one merely has superpositions.

*University of Namur, dept. of Business Administration, Rempart de la Vierge 8, 5000 Namur, Belgium

- recursive : the evolution operator ϕ generates a time series representing the evolution of the system over time. If **f** is the operator and the starting point is x_0 (to take a one-dimensional example), then $x_1 = f(x_0)$, $x_2 = f(x_1)$, $x_3 = f(x_2), ..., x_{n+1} = f(x_n)$. The n^{th} element, x_n then is $f^n(x_0)$[1].

Feigenbaum [FE80] also points out that this complex behavior is mainly generated by its recursive nature, rather than by the specific operation performed by the function.

There is very little agreement on how to actually detect the nature of chaos. One element is always considered to be the most characteristic, namely sensitivity to initial conditions [BDVT91]. In general, this means that the slightest difference in initial conditions of a dynamic system will cause the system to take a completely different orbit towards a potentially different final state. We will primarily focus on the sensitivity to initial conditions to describe the essence of chaos. We will also observe that this sensitivity to initial conditions covers a multitude of other phenomena for which we can compute a number of statistics. We furthermore limit ourselves to the period doubling bifurcation (which will be explained later) as a road to chaos. [2]

To determine whether or not there is sensitivity to initial conditions, the concept of strange attractors is very important. Attractors can be defined as a final state to which all trajectories converge. We might say that it constitutes the final solution of the non-linear differential equation in which the system will remain. We distinguish between regular attractors, corresponding with periodic and quasi-periodic solutions (or limit cycles) and irregular or strange attractors corresponding with aperiodic solutions. The nature of this attractor now determines whether or not there is sensitivity to initial conditions. The periodicity of the limit cycle represents the number of iterations the system needs to reproduce itself. Consequently, by observing the system for a certain lapse of time, we will find that the system behaves in a regular and orderly way. From any given state, we can consequently predict any future state of the system. What is now the implication of this with respect to the initial conditions ? When the attractor is periodic, the final state of the system will always be the attractor, irrespective of the initial conditions. Consequently, the information that is represented in the initial values of the system is destroyed as any nearby initial situation will always lead to the same final state.
When, on the contrary, the attractor is aperiodic, a totally different behavior of the system will be observed. Where in the periodic case, the system was perfectly predictable and orderly in its behavior, in the aperiodic case, the system behaves in an ostensibly random way. The aperiodicity of the attractor has as major implication that during whatever time length the system is observed, it will be virtually impossible to detect some kind of sequence that will permit to make any kind of prediction. In the periodic case, the system exhibits a high degree of self similarity : after a certain number of iterations (which depends on the periodicity), the system reproduces itself. In the aperiodic case, this self similarity diminishes and finally disappears after a number of iterations. This implies that it severely limits the possibility to predict from any given moment a future state of the system. Periodic attractors are also called simple attractors and aperiodic ones are called strange attractors [BT92].

4 Descriptive and Numerical Tools to study Chaos

Though by means of numerical experiments, the sensitivity to initial conditions can be revealed, there are a number of other tools at our disposal that allow to shed a more detailed light on the underlying chaotic process. These tools are presented in this section and will be explained from a 'user' point of view.

[1] $f^n(x_0)$ represents the n^{th} iteration of function f.

[2] Other, known and more or less understood ways to chaos are resonance overlap and intermittency [BV92].

4.1 Phase Space

One of the ways in which one may study the dynamic behavior of a system and detect the presence of attractors is by means of a phase space diagram which is a graphical representation of the evolution of a system towards a solution. Each point of that space then determines unequivocally the state of the system at a given time. The axes may represent the position and speed of the system and the corresponding space trajectory or orbit is the evolution of the system.

4.2 Embedding Dimension

When studying a dynamical system, we may not always have an idea on how the system looks like nor what variables influence its behavior. However, quite often one disposes of at least a one dimensional observation. Takens has developed a technique that allows to reconstruct the properties of the dynamical system under study solely on the basis of this one dimensional observation. We thus create an artificial dynamical system which, as Takens proved, has the same topological properties as the original system. This boils down to finding an appopriate dimension in which to represent the system, called the embedding dimension.

One proceeds in the following way. We assume that the time series is produced by a set of deterministic equations such as $x^i_{t+1} = f_i(x_t)$ where x $\in \Re^n$, i=1,...,n. As we said previously, we do not know the structural form of the system nor its true dimension (n). Nor are we certain that the time series contains the real x^i_t-values, so we represent the observed values by $\overline{x}^i_t$ and we assume that the following relation holds : $\overline{x}^i_t$=h(x_t) which means that the observed values in some way depend on the true value vector x^i_t.

We now take the last element of the time series and combine it with its **m** predecessors to form a vector of length m, $X^m_t = (\overline{x}^j_t, \overline{x}^j_{t-1}, ..., \overline{x}^j_{t-m+1})$. We do this for every element in the time series, eliminating of course the first m-1 elements. We thus obtain a series of m-dimensional vectors X^m_j. The length m is called the embedding dimension and each vector is called a m-history, describing a point in m-dimensional space. Takens now proved that this artificially reconstructed object is topologically equivalent to the real system if the following conditions hold :

- the variables x^i of the true system are located on the attractor.
- the functions$f_i(x)$ and $h_i(x)$ are smooth functions.
- m >2n-1.

This is a very important result because it allows to calculate three other very important measures (of which only the first one will be discussed here), namely the correlation dimension , the Kolmogorov Entropy and Lyapunov exponents.

In order to illustrate this somewhat abstract procedure, we apply it to the well known chaotic system, the Lorenz System which is composed of the following set of coupled non linear differential equations :

$$\begin{cases} \frac{dX}{dt} = a(Y - X) \\ \frac{dY}{dt} = bX - Y - XZ \\ \frac{dZ}{dt} = XY - cZ \end{cases}$$

For our purposes, we are not so much interested in the exact semantics of the variables but rather in the overall behavior of the system. We have computed 15.000 values (where dt=0.05) for X,Y and Z using the following parameter values : a=10, b=10 and c=-4. When reconstructing the attractor, we assume that we only dispose of a one dimensional observation of the system, nly. on the X-variable. To simplify things a bit and using the fact that the Lorenz System is of dimension 3, we fix the embedding dimension to 3.

In order to avoid autocorrelation (which would merely generate an attractor where all points are centered around the diagonal (plane)), we only use 1 out of every 5 observations. Some of the

Observation I	Observation II	Vector(1)	Vector(2)	Vector(3)
0.034168	0.347151	0.034168	0.087486	0.274364
0.038980	0.439782	0.898230	2.947274	...
0.046739	0.557647		...	...
0.057089	0.707582			
0.070430	**0.898230**			
0.087486	1.140478			
0.109239	1.447933			
0.136953	1.837410			
0.172245	2.329264			
0.217174	**2.947274**			
0.274364	3.717394			

Table 1: Creation of the Embedding Vector of Dimension 3

original values and their constructed vectors are given in Table 1. If we now plot both the original X,Y and Z-values (see Figure 1 (a)) and the artificially reconstructed system X,Y' and Z'-values (see Figure 1 (b)), we can indeed see that topologically the two are equivalent. [3]

4.3 Correlation Dimension

As we said previously, a fundamental characteristic of chaotic systems is its sensitivity to initial conditions which basically means that two orbits starting from very close initial points, become soon uncorrelated. However, this does not imply that the dynamical system can be on any position in phase space. Every orbit will always be located on an attractor and therefore all the points on the attractor are spatially related. Grassberger and Procaccia have developed a technique which measures the degree of spatial correlation between 2 points in phase space.

It is computed in the following way. The starting point is the artificial dynamical system as described in the previous section.. Two points $\overline{x}_t^i$ and $\overline{x}_t^j$ are said to be spatially correlated if the Euclidian distance is less than the radius **r** of an m-dimensional sphere centered on one of the two points, or put more formally :

$$||\overline{x}_t^i - \overline{x}_t^j|| < r$$

The spatial correlation between all points on the attractor is then measured by :

$$C(r,m) = \lim_{T\to\infty} \frac{1}{T^2} \sum_{i,j=1}^{T} H(r - ||\overline{x}_t^i \overline{x}_t^j||) \tag{1}$$

where T= length of the series of the constructed m-vectors x_t^m, $||.||$ is the Euclidean norm and

$$H = H(y) = \begin{cases} 1 & \text{if } y > 0 \\ 0 & \text{if otherwise} \end{cases}$$

C(r,m) is called the correlation integral from which we can compute the correlation dimension in the following way :

$$D^c(m) = \lim_{r\to\infty} \frac{ln[C(r,m)]}{ln[r]}$$ [4]

The value of C(r,m), and consequently of $D^c(m)$, depends on the length of the m-vector. It was shown, however, by Grassberger and Procaccia that for deterministic systems the correlation dimension converges to a fixed value, independently of the embedding dimension used for calculation.

[3] For generating these figures, we have used a time step of 1.

[4] It was also shown by Grassberger and Procaccia that $D^c(m)$ is a good approximation of the Hausdorff dimension ($D^c \leq D^H$), also known for its non-integer values as the fractal dimension.

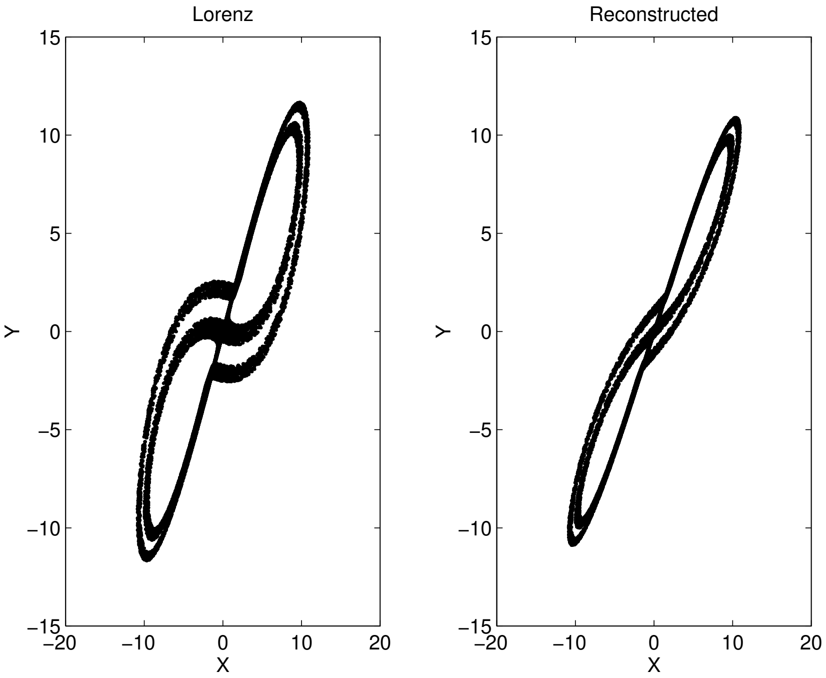

Figure 1: Reconstruction of the Lorenz Attractor

We therefore will compute the correlation integral C(r,m) for different embedding dimensions and we can then plot, on a log-log scale, the obtained values for C(r,m) given a particular radius.

We compute this for different radii until all points are enclosed by the hypersphere and we can plot this number as a function of the radius r. The slope of the log-log plot is then the correlation dimension. What we can observe is that the slopes of these lines do not change anymore, even though the embedding dimension changes. It is this 'constant' slope which yields the value of the $D^c(m)$. In the case of pure stochastic systems, the slope will permanently increase in function of the embedding dimension.

Figures 2 and 3 illustrate this for the Lorenz model, the SSSEM-model and for a detrended Standard & Poor 500 time series. As we can observe, a similar overall behavior can be seen.

4.4 Spectral Analysis

A well known numerical tool is spectral analysis which allows to analyse the periodicity of dynamic systems. The technique used is called the Fourier Transform.

When a time series does not converge, it can periodically oscillate, or may appear to be random. The Fourier analysis allows to distinguish between periodic or quasi-periodic series and random ones. However, we must be careful with this distinction. It may very well be that a time series appears random without being so. This may be caused by the fact that, for instance, the length of the observation interval is smaller than the periodicity of the system.

In the discrete case, we have two possible situations. The first one being either a periodic or quasi-periodic signal where we can observe important peaks in the power spectrum on the frequences of the periods, and eventually some smaller peaks, for the harmonic frequencies of the main frequences. The second possibility is the random case, in which the power spectrum seems continuous without special peaks. In that case, the power spectrum reflects broad band noise.

Figure 4 represents the fourier spectrum for two different time series. The first figure on the left contains the temporal evolution of the US$ interest rate and the figure on the left is the fourier

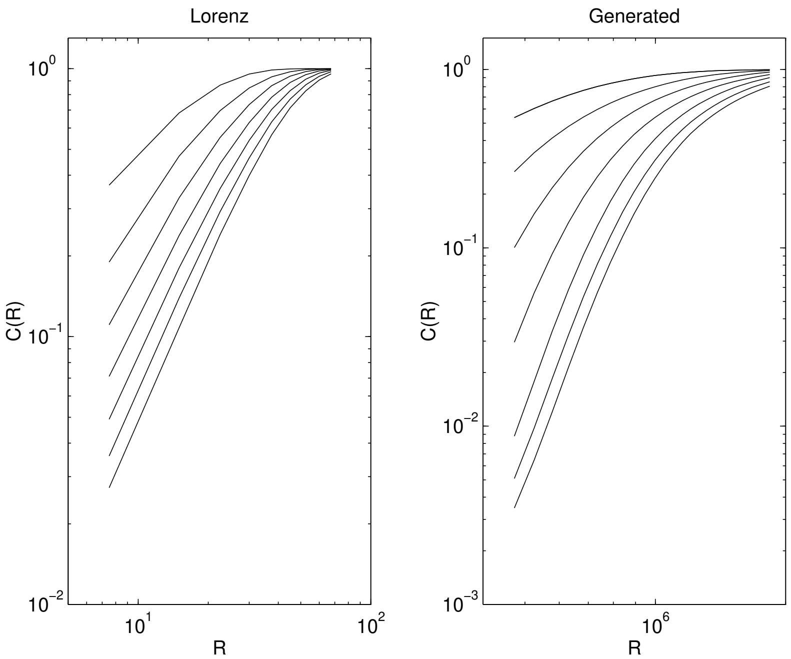

Figure 2: Function of the Correlation Integral C(r,m) and the radii r for the Lorenz Model and the SSSEM-model

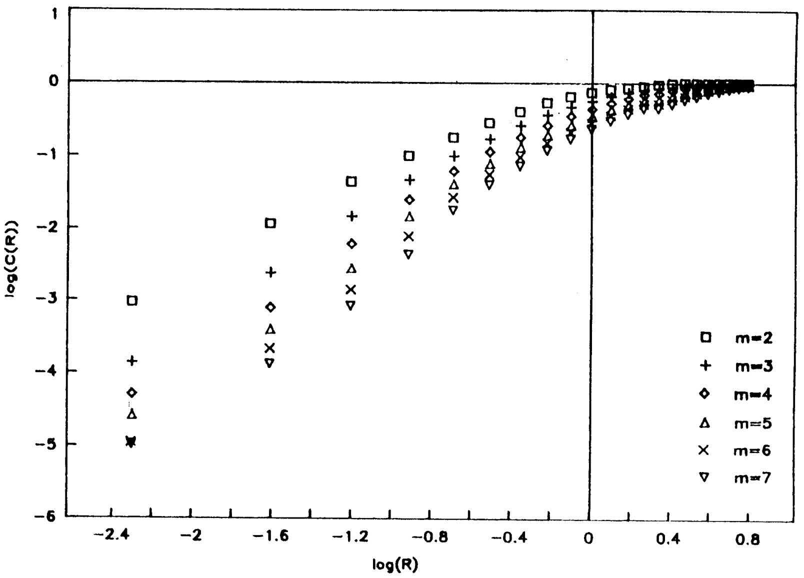

Figure 3: Function of the Correlation Integral C(r,m) and the radii r for S&P 500 [PE91]

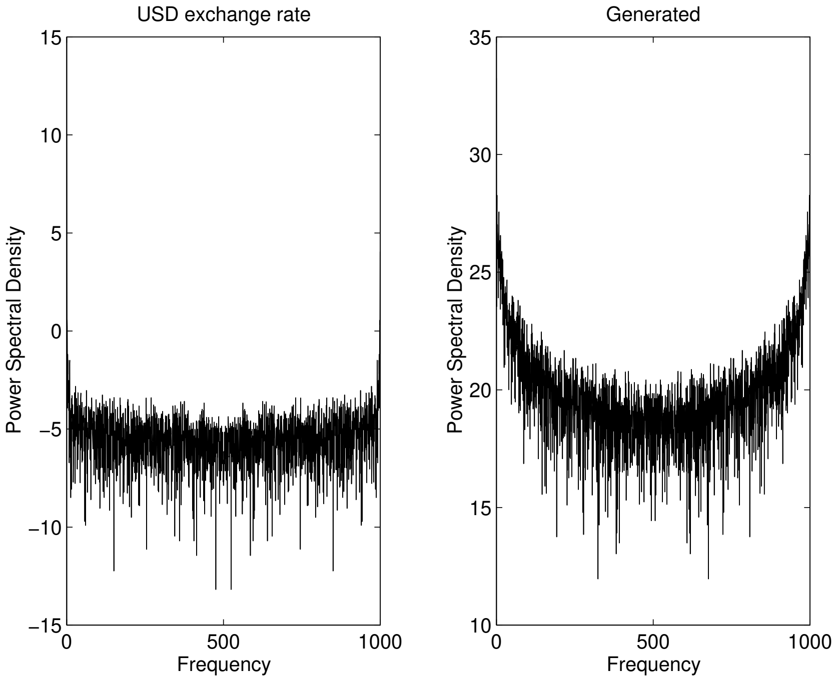

Figure 4: Fourier Spectrum for both real and generated chaotic time series

spectrum for the artificially generated data. In both cases, we observe the broad band spectrum indicating no periodic behavior.

5 Conclusion

When studying non linear and chaotic phenomena, we can not go very far analytically. We therefore have to use empirical studies, numerical experiments, econometric models or simulation models in order to be able to describe to a certain extent the behavior of the real world phenomena we are interested in. When restricting ourselves to the simulation technique, it is important to conserve the basic and fundamental characteristics of the system. If for instance the system has some chaotic properties, then the simulation model also must display this kind of behavior. Because chaos is not always easy to distinguish from aperiodic or even pure stochastic behavior, it is important to have tools that shed some light on this. We have not presented the whole array of numerical tools available but only those that are intuitively simple to understand. The simulation model of a financial market and agents acting on this market served as an example that artificially generated data indeed did have the same overall characteristics as the real world data.

References

[BDVT91] Broer H.W., Dumortier F., van Strien S.J., Takens F.; *Structures in Dynamics*, North-Holland, Amsterdam, 1991, 309 p.

[BT92] Broer, Takens; *Wegen naar Chaos en Vreemde Aantrekking*, in : [BV92], pp.1-76.

[BV92] Broer, Verhulst; *Dynamische systemen en chaos*, Utrecht, Epsilon Uitgaven, 1992, 349 p.

[DDE93] De Grauwe P., Dewachter H., Embrechts M.;*Exchange Rate Theory : chaotic models for foreign exchange markets*, Blackwell, Oxford, 1993, 273 p.

[FE80] Feigenbaum; *Universal behavior in nonlinear systems*, Los Alamos Science, 1980, pp.4-27.

[LO89] Lorenz H.; *Nonlinear Dynamical Economics and Chaotic Motion*, Springer Verlag, 1989, 248 p.

[PE91] Peters E.;*Chaos and Order in the Capital Markets*, John Wiley & Sons, New York, 1991, 240 p.