

## Detecting Nonlinear Dynamics Using BDS Test and Surrogate Data in Financial Time Series

Alex Neri Gutierrez<sup>1</sup>, Alexis Rodriguez Carranza<sup>2</sup> and Ana Gamarra Carrasco<sup>1</sup>

1. Department of Sciences, Private University Antenor Orrego, 13006, Peru

2. Department of Sciences, National University of the Trujillo, 13006, Peru

**Abstract:** Physicists experimentalists use many observations of a phenomenon, which are the unknown equations that describe it, in order to understand the dynamics and obtain information on their future behavior. In this article we study the possibility of reproducing the dynamics of the phenomenon using only a measurement scale. The Whitney immersion theorem ideas are presented and generalization of Sauer for fractal sets to rebuild the asymptotic behavior of the phenomena and to investigate evidence of nonlinear dynamics in the reproduced dynamics using the Brock, Dechert, Scheinkman test (BDS). The applications are made in the financial market which are only known stock prices.

Key words: Nonlinear dynamics, surrogate data, BDS test and serial times.

## **1. Introduction**

For a study of the asymptotic behavior of solutions of a system, the area of dynamical systems has developed a lot of tools, but in many phenomena as financial markets, equations that model them are unknown and the only available information is a temporal set of measures.

A time series is a function  $s: IR \rightarrow IR$ , and its image could be considered as observations taken overtime in a phenomenon. Those pieces of observations have information about the system, and we are ready to answer questions, like does that series have enough information to rebuild the dynamics? If it is so, can we predestine its future behavior? But there is one more general question. Will the financial market have a random behavior or not? The chaos theory could give us.

An answer where systems with an apparently random behavior, comes from a deterministic system, it means a completely modeling system by equations. A characteristic of these systems is the non-linear dependencies between their variables. We will do our research in a time series of the stock prices in the mining company MINSUR.SA to determine concrete evidence of non-linearity presence in that series. Fig. 1 shows the evolution of stock prices of the company.

The main task to study nonlinear behavior in a time series is the rebuilding of the dynamics it has. If a time series comes from a deterministic system, so that it has dependence between its components given by the system of equations that models and contains geometric information considering that trajectories converge to an attractor which will immerse in some Euclidian space. Using such information and the Whitney immersion theorem ideas in which a dimensional compact manifold n can be immersed in IR2n and its generalizations for fractal sets given for Tim Sauer [3], we will rebuild the dynamics knowing the information of one of the components of Lorenz system only.

Now we can see the probabilistic version of Theorem of Whitney [3].

## 2. Main Theorems

Theorem 1: (Prevalent theorem of Whitney). If A is a smooth compact manifold of dimension d contained on  $\mathbb{R}^k$ . Almost all smooth map, F:  $\mathbb{R}^k \rightarrow \mathbb{R}^{2d+1}$ , is an immersion of A.

**Corresponding author:** Alex Neri Gutierrez, msc, research field: dynamical system.



Fig. 1 Evolution of stock prices of the MINSUR company.

The theorem (1) says it is possible to rebuild the existing dynamics with projections in Euclidian spaces. Although we have the theorem of Whitney, we still have a practical problem. We only can get a temporal observation of the dynamics in the financial market; the evolution of the stock prices and we will need 2d + 1 observations of the phenomenon. So, the obtained results are not enough. Takens assumed this problem adding the contained dynamics in the time series in the Theorem of Whitney. It is a projection via a function of observation of the phase space where the dynamical system is developed for IR. Therefore, it contains information about the dynamics. For this, it defines the *delay coordinates*, which only need a temporary observation.

Definition 1: If  $\Phi$  is a dynamics system over a manifold A, T a positive integer (called of delay), and *h*:  $A \rightarrow IR$  a smooth function. We defines the map of delaying coordinates  $F_{(\Phi,T,h)}: A \rightarrow R^{n+1}$ 

$$F_{(\Phi,T,h)}(x) = (h(x), h(\Phi_T(x)), h(\Phi_{2T}(x)), \dots, h(\Phi_{nT}(x)))$$

Takens [4] demonstrated a new version of theorem of Whitney for the delaying coordinates.

Theorem 3: (Takens) A is a dimensional compact manifold *m* dimensional,  $\{\Phi_k\}_{k\in\mathbb{Z}}$  a discrete dynamical system over A, generated by  $F: A \to A$ , and a function of classes  $C^2$ ,  $h: A \to IR$ . Then a generic characteristic of map  $F_{(\Phi,h)}(x): A \to IR^{2m+1}$  defined by  $F(\Phi, h)(x) = (h(x), h(\Phi k(x)), h(\Phi 2k(x)), ..., h(\Phi nk(x)))$  is an immersion.

The final generalization used in this article was given by B. Hunt, T. Sauer and J. Yorke [3], that is a fractal version of theorem of Whitney for the delaying maps set with A being a fractal set.

Theorem 4 (Fractal Delay Embedding Prevalence Theorem)  $\Phi$  a dynamics system over an open subset Uof  $IR^k$ , and A which is a compact subset of U with box dimension d and n > 2d an integer and T > 0. Assume that A contains only a finite number of equilibria points, it does not contain periodic orbits of  $\Phi$  of period T o2T, a finite number of periodic orbits of  $\Phi$  of period 3T, 4T, ..., nT, and these periodic orbits of linearization have different eigen-values. So for almost every smooth function (in the sense prevalente) smooth function h over U, the delay coordinates map  $F_{(\Phi,T,h)}$ :  $U \rightarrow R^n$  is injective over A.

The theorem (4) does not provide an estimation about the smaller dimension for which almost every delaying coordinate map is injective. However, there are numerical algorithms which allow calculating the immersion dimension and the delaying time in the reconstruction. Following, we show the reconstruction of the dynamics generated for the system of Lorenz using only the coordinate x of the system.

# **3. Examples of Reconstruction of the Attractor Using Delay Coordinates**

We use the Lorenz's attractor to show the technique of delaying coordinates. The function of observation h was the projection in the x axis.

$$h: IR^3 \to IR$$
$$(x, y, z) \to h(x, y, z) = x$$

The time series will be formed by x coordinates of the trajectories that are numerical solutions of the equation. Fixed the dimension, n = 3, we change the value of T. According to Lorenz's we use delaying times, T = 1, T = 10, T = 17, T = 90. In the case T = 1points in the space are highly correlated and the graphic is almost a straight line. At the other extreme,



Fig. 2 Reconstruction of the Lorenz attractor using projections with time delay d = 1, 10, 17 and 90 respectively.

T = 90, points are not correlated and the gotten graphic does not represent the reconstruction of the attractor. The optimal delay time was T = 17.

## 4. Identifying Non-linearity

We will identify no lineal dependences on data. The most used techniques are method of surrogate's data and the Brock, Dechert, Scheinkman test (BDS).

Kantz e Schreiber [11] recommend that before doing any non-lineal analysis on a data set, it is good practice to check if there is no linearity. We presented the technique of surrogate data and the BDS test.

#### 4.1 Surrogate Data

A very useful technique used by physicist is surrogate generation of Surrogate data, a procedure done by Theiler, Eubank, Longtin, Galdrikan and Farmer [15]. From the original data, it generates a set of random series so that, these keep the linear properties of the original series (average, variance, Fourier spectrum) but eliminating the possible nonlinear dependencies. Then an indicator is evaluated, which is sensitive to nonlinear dependencies and tries to reject the null hypothesis, which states that data are obtained by a stochastic linear process. If the null hypothesis is true, then the substitute series procedure would not affect the indicator.

The most used indicators are correlation dimension. It was used by Small et al. [10] and no linear prediction was used by Kaplan [6].

We will illustrate the technique of substitute data applied in Lorenz series and use the nonlinear prediction.

The election was due to deterministic systems the prediction on a short term is possible, in contrast to Stochastic systems in which the prediction in a short term is impossible. For the prediction nonlinear we will use the method developed by Hegger, Kantz and Schreiber [5], where the value of predicting is the average of "future" values. This state n + k, is defined by average of the closed values. It is calculated by the following equation:



Fig. 3 In the graphic the projection errors of Lorenz series and its surrogate series.

$$\overline{S}_{n+k} = \frac{1}{|U_n|} \sum_{S_j \in U_n} S_{j+k},$$

where  $U_n$  is the neighborhood,  $S_j$  are values within  $U_n$ and  $S_{j+k}$  are the values at time n + k. With the prediction held, this is compared with the quadratic average error of prediction on future value of the original series with the surrogate series. If the prediction error of the series original is smaller than all the surrogate series, then we reject the null hypothesis. This graphic shows the results.

The projection errors of the original series were smaller than its substitutes. Like this, we reject the null hypothesis. The following technique is more about statistics.

## 5. BDS Test

It was developed by Brock, Dechert and Scheinkman [12]. It is based on the correlation dimension to detect nonlinear structure in a time series. Additionally, this test can be used to test how good the fit estimation model is.

Given a time series  $s_1, s_2, \ldots, s_N$ , using the method of delay coordinates, we have  $M = N - (m - 1)\tau$ vectors in  $IR^m$ ,  $X_i = (s_i, s_{i+\tau}, \ldots, s_{i+(m-1)\tau})$ , where  $\tau$  is the delay time, *m* is the immersion dimension. The correlation dimension is given by the formula:

$$C(M,r) = \frac{1}{M(M-1)} \sum_{i=1}^{M} M_{X_i^m}(\epsilon)$$
$$= \frac{1}{M(M-1)} \sum_{i \neq j} \theta(\epsilon - ||X_i - X_j||)$$

where  $M_{Xi}(\epsilon)$  is the function point mass, which indicates the probability where the points  $X_i$ ,  $X_j$  are  $\epsilon$ close to each other.

For practical and didactic issues, we consider  $\tau = 1$ . Now if m = 2 we will get:

$$X_i = (x_i, x_{i+1}),$$
  
 $X_j = (x_j, x_{j+1}).$ 

Then,  $||X_i - X_j|| \le E$  implies  $|x_i - x_j| \le \epsilon$  and  $|x_{i+1} - x_{j+1}| \le \epsilon$ .

This allows saying that if the points  $X_i$  and  $X_j$  are closing to each other, then, the points from series  $x_i$  and  $x_j$  as well. It happens the same with points  $x_{i+1}$  and  $x_{j+1}$ . Thus:

$$M_{X_i}(\epsilon) = \mathbb{P}(||X_i - X_j|| \le \epsilon)$$
  
=  $\mathbb{P}(|x_i - x_j| \le \epsilon; |x_{i+1} - x_{j+1}| \le \epsilon)$ 

where *P* indicates the probability. If the data  $x_1, x_2, ..., x_N$  are IID (independent and identically distributed), then  $P(|x_i - x_j| \le E) = P(|x_{i+1} - x_{j+1}| \le E)$  and we have:

$$M_{X_i}(\epsilon) = \mathbb{P}(|x_i - x_j| \le \epsilon) \mathbb{P}(|x_{i+1} - x_{j+1}| \le \epsilon)$$
  
=  $\mathbb{P}^2(|x_i - x_j| \le \epsilon).$ 

#### Detecting Nonlinear Dynamics Using BDS Test and Surrogate Data in Financial Time Series

Thus, in dimension m we have:

$$M_{X_i}(\epsilon) = \mathcal{P}^m(|x_i - x_j| \le \epsilon)$$

Then,

$$C_m(M,\epsilon) = \frac{1}{M} \sum_{i=1}^M M_{X_i}(\epsilon)$$
$$= \frac{1}{M} \sum_{i=1}^M P^m(|x_1 - x_2| \le \epsilon)$$
$$= P^m(|x_1 - x_2| \le \epsilon).$$

As  $P(|x_1 - x_2| \le \epsilon) = P(|x_1 - x_3| \le \epsilon) = \cdots = P(|x_i - x_j| \le \epsilon) = \cdots$  This suggests that the data are IID, then,

$$C_m(M,\epsilon) = C_1^m(M,\epsilon)$$

where  $C_m(M,\epsilon)$  is the integral of correlation in dimension *m* and  $C_1^m(M,\epsilon)$  is the integral of correlation in dimension one.

In Brock, Dechert and Scheinkman [12], statistical BDS test is defined by:

$$\sqrt{M} \frac{C_m(M,\epsilon) - C_1^m(M,\epsilon)}{\sigma_m(\epsilon)}$$

where  $\sigma_m(M, E)$  is the estimation of the standard asymptotic error:  $C_m(M, E) - C^m(M, E)$ . They proved that:

$$\sqrt{M} \frac{C_m(\epsilon) - C_1^m(\epsilon)}{\sigma_m(\epsilon)} \sim N(0, 1)$$

Now we will introduce an example applying the BDS test in Lorenz series and a series consisting of random numbers.

## 5.1 BDS Test for Lorenz Series

Table 1

Dimension fo immersion (m)	BDS
1	165.270466
2	447.623951
3	578.535241
4	772.617763
5	1078.467687
6	2338.679429
7	3593.653394
8	5640.618871

5.2	BDS	Test	for	a Serie	s of I	Random	Numl	bers

Table	2
Lanc	-

Tuble 2				
Dimension fo immersion (m)	BDS			
1	0.560594			
2	0.560594			
3	0.425293			
4	0.069765			
5	0.116118			
6	0.336857			
7	0.633835			
8	0.524584			

To understand the results, let us start talking about the degree of significance of a hypothesis. To ask a null hypothesis, in the case of the BDS test the null hypothesis is that the observations are IID, it can make a mistake of rejecting the null hypothesis being true. Suppose that the probability of committing this error is  $\alpha$ , namely:

 $\alpha = P$  (rejecting  $H_0|H_0$  is true)

 $\alpha$  is called the degree of significance of this mistake. Now, in the BDS test:

 $H_0: C_m(M, E) = C^m(M, E)$  $H_a: C_m(M, E) /= C^m(M, E).$ By the test BDS, we know that:

 $\sqrt{M} \frac{C_m(M,\epsilon) - C_1^m(M,\epsilon)}{\sigma_m(\epsilon)} \sim N(0,1)$ 

Then,

$$\alpha = P(\text{rejeitar} \ H_0 | H_0 \text{ is true})$$

$$\begin{split} \alpha &= P(|\sqrt{M}\frac{C_m(M,\epsilon) - C_1^m(M,\epsilon)}{\sigma_m(\epsilon)}| > z_c|H_0 \quad \text{is true} \ )\\ \alpha &= P(|\sqrt{M}\frac{C_m(M,\epsilon) - C_1^m(M,\epsilon)}{\sigma_m(\epsilon)}| > z_c \sim N(0,1))\\ \alpha &= 1 - 2P(0 \leq \sqrt{M}\frac{C_m(M,\epsilon) - C_1^m(M,\epsilon)}{\sigma_m(\epsilon)} \leq z_c) \end{split}$$

So,

$$P(0 \le \sqrt{M} \frac{C_m(M, \epsilon) - C_1^m(M, \epsilon)}{\sigma_m(\epsilon)} \le z_c) = \frac{1 - \alpha}{2}$$

If we look at the tables of the normal distribution N (0, 1), we see that for a degree of significance of  $\alpha = 5\%$ ,  $Z_c = 1.96$ . As, we reject the null hypothesis, with

degree of significance of 5%, if

$$\left|\sqrt{M}\frac{C_m(\epsilon) - C_{1m}(\epsilon)}{\sigma_m(\epsilon)}\right| > 1.96$$

Noting again Tables 1 and 2, we can observe that in the case of Lorenz series, we reject the null hypothesis, the data are IID, otherwise case series formed by random numbers.

## 6. Results

The time series studied was the series of prices of the mining company Minsur, counting with the prices from 10/13/1993 to 11/26/2014, with a total of 21 years of daily observations.

For the use of surrogates, substitute series are generated that preserve linear properties of the original series, for example, Fig. 4 shows the histograms of the original series as well as their substitute series.

The next step is the calculation of the predictions of

each of the series, original and substitutes, comparing the relative errors. In Fig. 4 it is observed that the prediction errors of the series, the original presents a relatively short prediction error in the short term and the substitute series shows a high prediction error from the start.

Therefore, the null hypothesis ( $H_0$ : Error of the original series = than the prediction error of all its substitute series) is rejected since the prediction error of the original series is smaller than the prediction error of all its substitute series. Therefore, we reject the fact that the original series is generated by a stochastic linear process.

On the other hand, the BDS test is a more statistical alternative to detect non-linear dependencies, we perform the test for several immersion dimensions. The next table shows the BDS statistic values for dimensions from 1 to 15.



Fig. 4 Histograms of the original series as well as their substitute series for Minsur prices.



Fig. 5 Error of the original series is smaller than the prediction error of all its substitute series.

Immersion dimension (m)	BDS
1	165.270466
2	195.653634
3	236.554950
4	296.226274
5	382.825922
6	508.150676
7	690.060422
8	955.415301
9	1,344.764026
10	1,919.802065
11	2,774.198843
12	4,050.599850
13	5,967.524832
14	8,860.578053
15	13,419.054424

We see in the table that | BDS | > 1.96. Therefore, we reject the null hypothesis (that the data are IID), with a degree of significance of 5%.

## 7. Conclusions

Both tests showed that the data contain non-linear dependences. Having non-linear dependence on the data, a work to follow to understand the complexity of the dynamics is the search for the presence of unstable periodic orbits, using for this, the maps of recurrence or maps of Poincare, as well as the presence of a fractal attractor.

## Acknowledgments

The authors thank the Department of Sciences of the Private University Antenor Orrego, UPAO, where the present work was developed. This project was financed by FAIN 2018 project-UPAO.

## References

- [1] Kantz, H., and Schreiber, T. 2004. *Nonlinear Time Series Analysis*. Cambridge University Press.
- [2] Mañe, R. "On the Dimension of the Compact Invariant Sets of Certain Nonlinear Maps. Dynamical Systems and Turbulence." *Warwick, Lecture Notes in Mathematics* 898: 230-42.
- [3] Sauer, T., Yorke, J. A., and Casdagli, M. "Embedology." Journal of Statistical Physics 65: 579-616.
- [4] Takens, F. 1981. Detecting Strange Attractors in Turbulence. Springer Berlin, Book Series Lecture Notes in Mathematics, 366-81.
- [5] Hegger, R., Kantz, H., and Schreiber, T. 1999. "Practical Implementation of Nonlinear Time Series Methods: The TISEAN package." *Chaos* 9: 413; doi:10.1063/1.166424.
- [6] Kaplan, D. 2002. "Making and Using Surrogate Data." http://www.macalester.edu/~kaplan/barcelona/surrogatew rap.pdf.
- [7] Fraser, A. M., and Swinney, H. L. 1986. "Independent Coordinates for Strange Attractors from Mutual Information." *Physical Review A (General Physics)* 33 (2): 1134-40.
- [8] Provenzale, A., Smith, L. A., Vio, R., and Murante G.

1992. "Distinguishing between Low-Dimensional Dynamics and Randomness in Measured Time Series." *Physica D* 58: 31.

- [9] Small, M., Yu, D., Simonotto, J., Harrison, R. G., Grubb, N., and Fox, K. A. A. 2002. "Uncovering Non-linear Structure in Human ECG Recordings." *Chaos, Solitons, Fractals* 13 (8): 1755-62.
- [10] Schreiber, T., and Schmitz, A. 1999. "Improved Surrogate Data for Nonlinearity Tests." *Phys. Rev. Lett.* 77: 635.
- [11] Kantz, H., and Schreiber, T. 1997. *Nonlinear Time Series Analysis*. Cambridge University Press.
- [12] Broock, W. A., Scheinkman, J. A., Dechert, W. D., and LeBaron, B. 1996. "A Test for Independence Based on the Correlation Dimension." *Journals in Its Journal*

Econometric Reviews 15 (3): 197-235.

- [13] Scheinkman, J. A. 1990. "Nonlinearities in Economic Dynamics." *Royal Economic Society, Economic Journal* 400: 33-48.
- [14] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. 1992. "Testing for Nonlinearity in Time Series: The Method of Surrogate Data." *Physical D: Nonlinear Phenomena* 58 (1-4): 77-94.
- [15] Hsieh, D. A. 1991. "Chaos and Nonlinear Dynamics: Application to Financial Markets." *The Journal of Finance* 46 (5): 1839-77.
- [16] Willey, T. 1992. "Testing for Nonlinear Dependence in Daily Stock Indices." *Journal of Economics and Business* 44 (1): 63-76.