# NON-LINEAR NOISE REDUCTION AND DETECTING CHAOS: SOME EVIDENCE FROM THE S&P COMPOSITE PRICE INDEX

Robert G. Harrison (1), Dejin Yu (1), Les Oxley (2), and Weiping Lu (1)
(1) Department of Physics, Heriot-Watt University, Scotland
(2) Department of Economics, University of Waikato, New Zealand

**Abstract:** Academic and applied researchers in economics have, in the last ten years, become increasingly interested in the topic of chaotic dynamics. In this paper we undertake nonlinear dynamical analysis of one representative time series taken from financial markets, namely the Standard and Poor's (S&P), Composite Price Index. The data is based upon (adjusted), daily data from 1928-1987, comprising 16127 observations. The results in the paper, based on the Grassberger-Procaccia (GP) correlation dimension measurement in conjunction with nonlinear noise filtering and the surrogate technique, show strong evidence of chaos in one of these series, the S&P 500. The analysis shows that the accuracy of results improves with the increase in the number of recording points and the length of the time series, 5000 data points being sufficient to identify deterministic dynamics.

## 1.0 Introduction

Chaos is widely found in the fields of physics and other natural sciences, however, the existence of chaos in economic data is still an open question. Various contributions have been made to this economic literature including, Barnett et al. (1994), Barnett and Chen (1988), Chen (1996), Brock and Sayers (1989), and Ramsey, Sayers and Rothman (1990). In addition, a new international journal which is exclusively devoted to, and entitled, *Studies in Nonlinear Dynamics and Econometrics,* has recently been founded which is testimony to the interest in this area.

Some of the main problems which pervade the area of economic time series evidence on chaos are the effects of noise, trend, and more general structural change. Of these noise and time evolution appear to be the most problematic with the latter often modelled via ARCH/GARCH processes which allow changing means and variances. These problems are compounded by the often paucity of the available data. In attempting to answer questions relating to the existence of nonlinearities and chaos in economic data, researchers have normally used either the Hinich bispectrum test, the BDS test of Brock, Dechert and Scheinkman (1996), White's (1989), test or more recently Kaplan's test, to identify nonlinearities and when considering chaos have used tools based on phase space reconstruction developed and used successfully in the physical sciences. The most commonly used of these chaos tests are the Lyapunov exponents test and the Grassberger-Procaccia (GP) correlation dimension test. While these tests have revealed an abundance of previously unexplained nonlinear structure and yielded a deeper understanding of the dynamics of many different economic time series, the case of deterministic chaos in these types of series is yet to be clarified. The main problem we identify in this study is that of noise which degrades these measurement techniques. The use of conventional filtering methods such as low pass filtering using Fourier transforms, moving averages etc., and also singular spectrum analysis based on singular value decomposition commonly used in economics, can lead to distortion of the dynamics.

## 2.0 Testing approach

Before economic data can be analysed for the existence of deterministic chaos, the twin problems of growing time trends and noise require consideration. The main contribution of this paper will be to the latter where new nonlinear noise reduction (NNR), techniques will be applied to the data. However, the following general methodology will be followed.

Firstly, the (log) data will be adjusted to remove systematic calendar effects and trend effects will be removed by differencing following Nelson and Plosser (1982). Currently, application of the Hodrick-Prescott (1981), to the data is being undertaken to assess the sensitivity of such filters.

Secondly, in order to reconstruct a chaotic attractor in phase space, two basic paramters, the embedding dimension *m,* and delay time *h,* must be correctly determined. The embedding theorem (*m=2d+1*, where *d* is correlation dimension), of Mane and Takens provides a sufficient condition in reconstructing an attractor from a scalar time series. An efficient method to determine an acceptable minimum *m*, from experimental time series is the so-called *false nearest neighbor* (FNN), recently developed using a geometrical construction. It monitors the behaviour of near neighbours under changes in the embedding dimension from $m \to m+1$. When the number of the false nearest neighbours arising through projection is zero in dimension *m*, the attractor has unfolded in this embedding dimension m. This technique is robust to the noise and a correct region of the embedding dimension can be determined in the presence of noise, which is important for the type of data used here. An estimate of the value of the delay time *h*, is provided by the autocorrelation function (ACF).

The Lyapunov exponents test and the Grassberger-Procassia correlation dimension method are well documented methods used in the quantitative analysis of time series data as tests for chaos, see for example, Abarbanel et al. (1993). Here we concentrate on the latter.

The geometrical features of an attractor can be specified using the Grassberger-Procaccia *correlation dimension.* Suppose we have a scalar time series $x_i\ (i = 1,2,\ldots N)$ of a dynamical variable sampled at an equal time interval $\Delta t$ from which the *K* vectors $\mathbf{Y}_j$ (*j=1, 2,...K*) in the *m*-dimensional phase space can be reconstructed using the time delay technique. Then the *correlation dimension* $D_2$ is defined and calculated as:

$$D_2 = \lim_{\varepsilon \to 0} \frac{\log_2 C_m(\varepsilon)}{\log_2 \varepsilon}, \qquad (1)$$

where $C_m(\varepsilon)$ is known as the correlation integral and can be computed as

$$C_m(\varepsilon) = \lim_{K \to \infty} \frac{1}{K(K-1)} \sum_{ij}^{K} \theta(\varepsilon - \|\mathbf{Y}_i - \mathbf{Y}_j\|),$$

(2)

where $\theta(x)$ is the Heaviside step function and $\|\mathbf{Y}_i - \mathbf{Y}_j\|$ is the distance between the vectors $\mathbf{Y}_i$ and $\mathbf{Y}_j$.

Thus, the sum $\sum_{ij}^{K} \theta(\varepsilon - \|\mathbf{Y}_i - \mathbf{Y}_j\|)$ is equal to the number of pairs *(i,j)* whose distance $\|\mathbf{Y}_i - \mathbf{Y}_j\|$ in the reconstructed phase space is less than the distance ε. For a chaotic attractor, $D_2$ is a non-integer, the value of which determines whether the system is low- or high-dimensional.

The use of this approach must, however, be applied with caution since it describes a kind of scaling of behaviour in the limit as the distance between points on the attractor approaches zero and therefore is sensitive to the presence of noise. Indeed our numerical experiments have shown that a noise level as small as 2~5% of the time series content can make these measurements inaccurate and inconclusive. Moreover, noise is also a main source to prevent precise predicition. Here we use NNR algorithms based on finding and extracting the approximate trajectory which is close to the original clean dynamics in reconstructed phase space from the observed time series. The implementation of the algorithms involves three basic steps: i) to reconstruct the underlying attractor from the observed series, ii) to estimate the local dynamical behaviour choosing a class of models and fitting the parameters statistically, and iii) to adjust the observations to make them consistent with the clean dynamics. The technique can reduce noise by about one order of magnitude. If some standard techniques are employed to pre-process the data, such as band-pass filtering, filtered embedding and singular value decomposition, significantly larger amounts of noise can be reduced since the local dynamics are largely enhanced.

These nonlinear noise reduction algorithms have been develped under the assumption that the noise is additive rather than dynamic. In practice, the two may not be distinguishable, being based on data only and both of them can be reduced as long as the exact dynamics can be reconstructed. However, in the following cases, the algorithms will fail: i) where the data are purely stochastic, ii) the deterministic dynamics are so weak that the data are uncorrelated, iii) the dynamic noise has driven the trajectories far from the exact dynamics of the nonlinear system, which becomes ashadowing problem. These methods are applied to the data prior to the application of the quantiative measurements for chaos.

While the correlation dimension measurement is often accepted as 'proof' of chaos, it is not a definitive test against time series data with certain types of coloured noise. This issue will be resolved using *surrogate* techniques. The correlation dimension method must be applied together with a surrogate technique to reliably discriminate between chaos and noise from a time series so as to avoid claims of chaos when simpler (such as linearly correlated noise), can explain the data. The consideration of surrogate data is based on the following. 1) Statement of a null hypothesis that shall be tested for consistency with the recorded original data, 2) Generation of a number of *surrogate* data sets, an algorithm is to randomise the phase of the raw data so that the surrogated set has the same Fourier spectra as the original, 3) Calculation of the value of interest, e.g., correlation dimension, Lyapunov exponents etc., for the original and all the surrogates, 4) Calculate mean and spread of the results obtained from the surrogates to determine whether the difference to the original, if any, is statistically significant.

## 3.0 Some results based upon the Standard and Poor's Composite Price Index

Although several economic data sets have been considered, we will report only one based upon the Standard and Poor's Composite Price Index which comprises 16127 daily observations on the logarithmic price change $x_t = 100[\log(p_t) - \log(p_{t-1})]$. For details of the data set see Gallant et al. (1993). The time series $x_t$ has been adjusted to remove systematic calendar and trend effects and is taken to be jointly stationary. A representative window of the raw series taken after 1947 is shown as Figure 1(a). Our analysis, based upon the GP correlation dimension measurement, in conjunction with the nonlinear noise reduction filtering and surrogate technique provides strong evidence in favour of chaos in this data. Some results are detailed below. As shown in Figure 2(c) (open squares), the raw data gave no saturated correlation dimension on increasing embedding dimension, suggesting that the data may be noise-dominated. Figure 1(b) presents the data after NNR of the raw data. Comparison of this with the raw data shows that the effect of noise is manifested as relatively large amplitude random fluctuations masking the overall deterministic patterns, in which the noise level is estimated to be ~90% of the clean signal. Correlation dimension analysis of this noise-filtered data revealed a clear (scaling) saturation region on increasing the embedding dimension as shown in Figure 2 (solid circles in (c)), indicative of deterministic chaos.

In confirming that the convergent correlation dimension is a result of chaotic dynamics, both the raw and noise-filtered series were surrogated to randomise the phase and so destroy the deterministic structure. The results show the dimension to diverge as shown in Figure 2(c) (open circles), consistent with stochastic behaviour, from which we confirm that the saturated correlation dimension of the noise-filtered data in Figure 2(c) arises from an endogenous deterministic mechanism. The accuracy of these results improves with the increase in the number of recording points and the length of the time series; 5000 data points being sufficient to identify deterministic dynamics. A particularly interesting finding is the overriding prevalence of stochastic behaviour in the first 5000 data points corresponding to the period 1928-1946, beyond which the data, analysed in either sets of 5000 data points or as a whole, displayed the deterministic chaotic dynamics with a correlation dimension $D_2 \sim 4.5$. These characterisations may be a consequence of the economic and social shocks caused by the Great Depression and subsequent Second World War.

## 4.0 Conclusions

Using the Standard and Poor's Composite Price Index comprising 16,127 daily observations from 1928-1987, we have identified the existence of deterministic chaos. Furthermore, the pre-WW2 period appears to be more volatile than the post-War period. Nonlinear noise reduction methods appear crucial in the removal of noise and this may transpire to be a general issue in the identification of chaotic dynamics in economic data.

In future work, we will be considering the implications of different transformations (for example Hodrick-Prescott) and different versions of nonlinear noise reduction filters. Furthermore, we will be considering forecasting and control, again emphasising nonlinear aspects.

## 5.0 Acknowledgements

We wish to thank Professor Ron Gallant for providing the data used in the paper.

## 6.0 References:

Abarbanel, H., R. Brown, J. Sidorowich, and L. Tsimring, The analysis of observed chaotic dynamics in physical systems, *Review of Modern Physics,* 65, 1331-92, 1993.

Barnett, W., and P. Chen, The aggregation-theoretic monetary aggregates are chaotic and have strange attractors: An econometric application to mathematical chaos, in W. Barnett, E. Berndt, and H. White, (eds.), *Dynamic Econometric Modelling.* Cambridge University Press, Cambridge, 1988.

Barnett, W., R. Gallant, M. Hinich, J. Jungeilges, D. Kaplan, and M. Jensen, A single-blind controlled competition among tests for nonlinearity and chaos, *Working Paper 190,* Washington University, St. Lois, Missouri, 1994.

Brock, W., W. Dechert, and J. Scheinkman, A test for independence based on the correlation dimension, *Econometric Reviews,* 15, 197-235, 1996.

Brock, W., and C. Sayers, Is the business cycle characterised by deterministic chaos? *Journal of Monetary Economics,* 22, 71-80.

Chen, P., Study of chaotic dynamical systems via time-frequency analysis, *Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis.* Piscataway, New Jersey: IEEE, 357-60, 1993.

Chen, P., A random walk or color chaos on the stock market? Time frequency analysis of S&P index, *Studies in Nonlinear Dynamics and Econometrics,* 1, 87-123, 1996.

Hodrick, R., and E. Prescott, Post-War US business cycles: An empirical investigation, *Discussion Paper 451,* Carnegie-Mellon University, 1981.

Nelson, C., and C. Plosser, Trends and random walks in macroeconomic time series, some evidence and implications, *Journal of Monetary Economics,* 10, 139-62, 1982.

Ramsey, J., C. Sayers, and P. Rothman, The statistical properties of dimension calculations using small data sets: Some economic applications, *International Economic Review,* 31, 991-1020, 1990.

White, H., An additional hidden unit test for neglected nonlinearity in multilayer feed forward networks, *Proccedings of the International Joint Conference on Neural Networks,* Washington DC, vol. 2, 451-5, 1989.

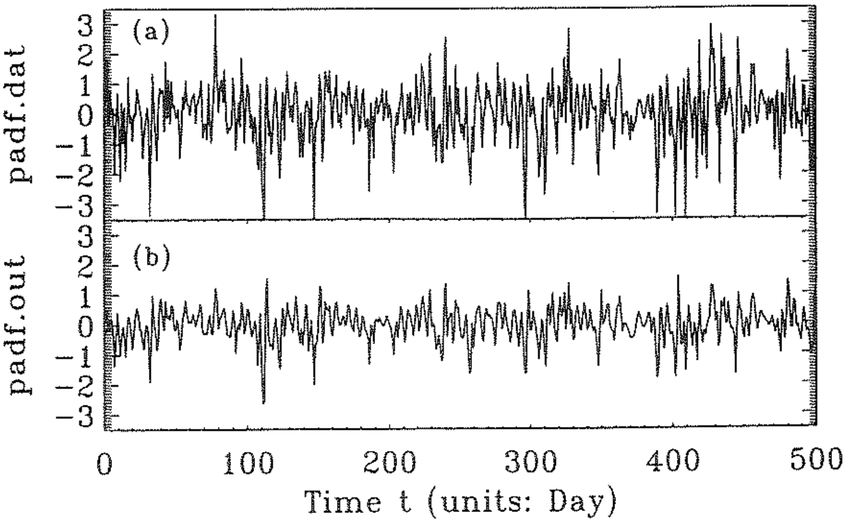


Figure 1. Two time series of $x(t)$ (a) raw data and (b) data after NNR.

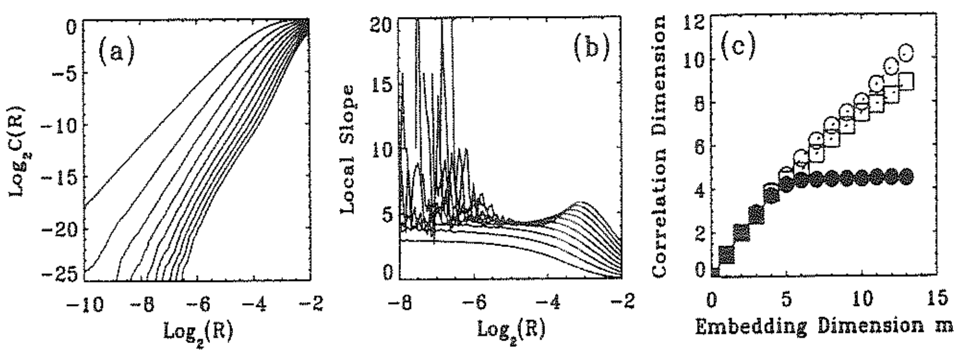


Figure 2. (a) Logarithmic plot of correlation function $C_m(R)$ versus correlation distance $R$ for embedding dimension $m = 3, 4, ..., 13$, (b) Local slope as a function of $\text{Log}_2(R)$ derived from (a), and (c) Correlation dimension $D_2$ versus embedding dimension $m$: □ - data before NNR, ● - data after NNR, and ○ - surrogate set from the data after NNR.