DATA MANIPULATION, CHAOS AND STOCK PRICES

Rae Weston [1)] James Wang [2)]

1) Professor of Management at Macquarie Graduate School of Management, N.S.W. Australia (rweston@laurel.ocs.mq.edu.au)
2) Ph.D. student of Management at Macquarie Graduate School of Management, N.S.W. Australia (jwang@bigpond.net.au).

Abstract

In this paper we explore the significance of a number of data manipulation techniques with respect to the identification of chaotic structures for the daily share prices of three leading Australian companies over the period November 1989 to November 1999. The companies are BHP, News Corporation and Southcorp Ltd.

We test for chaotic structures using original data, first difference, logs, trend-deduction and other filters including. We are able to establish that the choice of data manipulation technique is crucial to the identification of chaotic structures, which we define as a non-integer fractal dimension and a positive Lyapunov exponent.

Using the best results from our test we also examine the accuracy of short-term forecasting and accuracy decay.

1. Introduction

1.1 Identification of chaos in time series

Chaos theory is an attractive process for understanding the non-linearity of time series, which may be useful in modeling financial markets now that it is recognised that market information may be less than complete for all participants. The identification of a chaotic structure may be achieved by three specific measures. First, by examining the Lyapunov exponent, which measures how nearby data evolves within a phase space of given dimensions. In a one-dimensional system the Lyapunov dimension is equivalent to the rate at which the system provides information, and, even in multi-dimensional cases the information rate and also the accuracy of a prediction by the system is dominated by the largest Lyapunov exponent. Provided one of the Lyapunov exponents is positive, that is the nearby data will move away from each other, this satisfies the typical behaviour expected of a chaotic structure.

The second measure is the correlation dimension, which seeks to identify the fractal dimension geometrically. This measure is particularly sensitive to the data set. It takes a hyperdimensional sphere of embedding dimension and radius and then calculates the fraction of subsequent data points within that sphere for various values of the radius. The correlation dimension taken from the plot of the log of the above fraction against the log of the radius, is the average slope of the cumulative curve over the middle quarter of the vertical scale, with the error taken to be half the difference of the maximum and minimum slopes over the same range. As the embedding dimension increases the correlation dimension also increases but saturates at the correct level. A low correlation dimension of less than 5 satisfies an expectation that the structure may be chaotic.

Third, as a non-linear prediction method singular value decomposition (Rowlands and Sprott,1992) can be used to reconstruct the phase-space of the underlying chaotic data structure and to predict short-term values.

1.2 Australian companies tested

This paper investigates two interesting aspects of the application of chaos theory to the identification of chaotic structures in the Australian share market: the sensitivity to the initial condition and the required manipulation of data and considers them with respect to the description of the structure of the series of three stock prices: Broken Hill Pty Ltd (BHP), News Corporation (NCP) and Southcorp Ltd (SRP) for daily close prices over the period 1 November 1989 to 4 November 1999. We incorporate in this study an insight supplied by Urbach (2000) who distinguishes between additive noise and dynamical noise. It will be remembered that noise disturbs our ability to identify the underlying system. Additive noise exists when either the environment obscures the true state of the system from measurement, or the device with which the system is observed is inaccurate, while dynamical noise enters the system either by altering the system's dynamics or through a perturbation of state.

2. Methods of data manipulation

In order to identify chaotic structure, we wish to eliminate noise as far as possible without compromising the underlying structure. From an earlier study (Wang and Weston,1999), we recognised the significance of data manipulation in the search for a suitable underlying chaotic structure and in the present paper, we explore a range of data manipulation methods, as summarised in Table 1.

Table 1 Data Manipulation Methods Used

No	Method	Manipulation
1.	Original data	Nil
2.	First difference	Today's price minus yesterday's
3.	Detrended original	Formula (1.1)
4.	Natural logs	Applying natural logs to 1.
5.	Detrended	Applying formula (1.2)

Methods 1,2 and 4 are adequately described in Table 1.

The third method, Detrended original involves removing the linear average growth from the original data, with the formula:

$$D_i = P_i - (s * i + \text{constant}) \qquad (1.1)$$

Where D_i = detrended original price series
P_i = original price series
i = observation number
s = average growth rate (slop) from the first data to the last data

The fifth method Detrended applies the formula used by Chen (1988):

$$S_i = \text{Log}_e(P_i) - (a * i + \text{constant}) \qquad (1.2)$$

Where S_i = detrended price series
P_i = original price series
i = observation number

Chen's formula uses the log scale in order to rationalise the magnitude of price changes, which tend to be larger over time, and rotates the x-axis back to level by using the slope, a, multipled by the observation number, i, in order to deal with non-stationarity in the data.

3. Test results

The test results of our search for chaotic structures from the three companies with five different data munipulation methods are summarised in Tables 2 to 6. In the tables, the data series will be classified as chaotic, if the value of Correlation Dimension is saturated and less than 5, and the largest Lypunov exponent is positive. The next prediction hit rate (N. Pre. Hit R.) is defined as the next prediction value divided by the next real value. It is good prediction, if it is positive and within 70% of accuracy.

Table 2. The Original data series

Stock	Cor. Dim. /Saturation	Largest Lypunov Exp	Chaotic	N. Pre. Hit R.
BHP	4.245 ± 0.398/yes	0.194 ± 0.020	Yes	99.8%/good
NCP	3.264 ± 0.574/yes	0.119 ± 0.025	Yes	100.7%/good
SRP	4.530 ± 4.530/no	0.143 ± 0.021	No	99.5%/good

Table 3. The Difference data series

Stock	Cor. Dim. /Saturation	Largest Lypunov Exp	Chaotic	N. Pre. Hit R.
BHP	6.020 ± 6.020/no	0.179 ± 0.016	No	-109.6%/bad
NCP	5.566 ± 5.566/no	0.134 ± 0.017	No	125.5%/good
SRP	5.988 ± 5.988/no	0.152 ± 0.016	No	76.5%/good

Table 4. The Detrended Original data series

Stock	Cor. Dim. /Saturation	Largest Lypunov Exp	Chaotic	N. Pre. Hit R.
BHP	4.738 ± 0.512/no	0.183 ± 0.020	No	83.0%/good
NCP	4.452 ± 0.353/no	0.168 ± 0.022	No	121.1%/good
SRP	4.721 ± 4.721/no	0.195 ± 0.021	No	257.9%/fair

Table 5. The Natural Logarithms data series

Stock	Cor. Dim. /Saturation	Largest Lypunov Exp	Chaotic	N. Pre. Hit R.
BHP	4.885 ± 0.608/yes	0.182 ± 0.020	Yes	100.1%/good
NCP	4.136 ± 1.195/yes	0.080 ± 0.021	Yes	100.1%/good
SRP	4.449 ± 0.395/no	0.146 ± 0.020	No	100.0%/good

Table 6. The Detrended data series

Stock	Cor. Dim. /Saturation	Largest Lypunov Exp	Chaotic	N. Pre. Hit R.
BHP	5.156 ± 0.574/no	0.161 ± 0.018	No	199.6%/fair
NCP	4.536 ± 4.536/no	0.125 ± 0.021	No	135.8%/good
SRP	4.424 ± 4.424/no	0.229 ± 0.022	No	284.1%/fair

From the above results, we find:

BHP and NCP in both Original and Natural Logarithms data forms can be classified as chaotic dynamics. All the others fail to show the existence of a low (less than 5) correlation dimension and can not be classified as chaotic dynamics.
Except for the prediction for BHP in the Difference form, all other predictions for next value are good or fair. This implies that a certain predictability may exist even where the data is not a true chaotic dynamic. However, the

validity of the prediction requires more vigorous evaluation through further study by comparing a longer period of true value and predictions.

Although BHP and NCP satisfy the chaotic dynamic criteria under Original and Natural Logarithms forms, both BHP and NCP demonstrate a non-stationary growth, which may interfere with the system dynamics and should be removed in the search of a more accurate chaotic dynamics. This can be seen from Fig 1. in the calculation of the Correlation Dimension of NCP, which exists a wider range of variance than in a typical Choas data series.

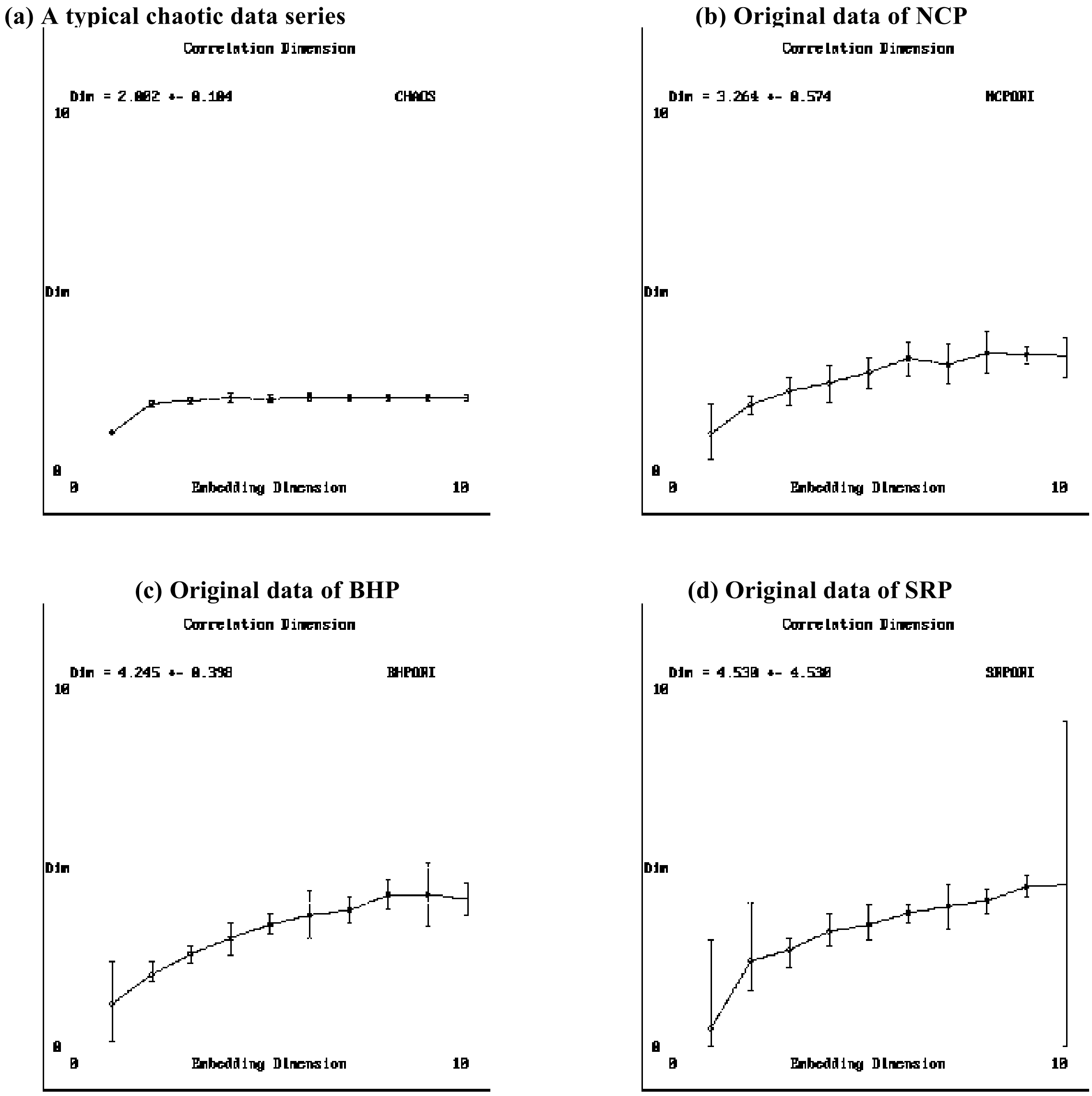

Figure 1. The comparison of the calculation of Correlation Dimension of a typical chaotic data series with BHP, NCP, and SRP stock prices. (a) A typical chaotic data series shows the correlation dimension saturated quickly with the increasing of the embedding dimensions, and with small variance in the value of correlation dimension. (b) and (c) show the correlation dimensions of BHP and NCP saturated with higher variance. It implies that the noise level within the chaotic dynamics is higher than in the typical chaotic model. (d) shows that SRP fails to generate a correlation dimension and is rejected as a candidate of a chaotic dynamics.

Comparing to Wang and Weston (1999), where it was shown that Detrended NCP had a chaotic dynamics with a different test time period (29 January 1988 to 31 December 1997), it implies that the choice of linear growth factor (a in formula 1.2) may alter the underlying dynamics of the system.

4. Prediction accuracy decay

One of the characteristics of chaos structure is the decay in predictive accuracy as time increase, which is estimated through the Lyapunov exponents. We use the four chaotic models identified to predict the next five periods and compare these to the actual value to see whether the prediction accuracy decay or not. The result is as table 7.

Table 7. The accuracy decay of chaotic dynamics

Stock	1st Pred.	2nd Pred.	3rd Pred.	4th Pred.	5th Pred.	Acc. Decay
BHP Orig.	99.8%	101.3%	100.2%	99.2%	96.9%	Well
NCP Orig.	100.1%	99.8%	99.2%	98.8%	98.4%	Well
BHP Ln	100.7%	100.7%	99.2%	93.7%	92.8%	Well
NCP Ln	100.1%	100.1%	100.0%	99.1%	98.8%	Fairly

From the Table 7, it is clear the predictive accuracy does decay. The predictive accuracy percentage seems very high. However, it is not very practical in reality since the share price of daily change is a very small percentage as well.

5. Conclusion

In this paper, we are able to identify chaotic structures from the Original and Natural Logarithm forms of BHP and NCP. However, certain variances that exist during the calculation of Correlation Dimension imply there is still a noise level interfering with the underlying chaotic structures. Further investigation of other techniques of data manipulation may still required to removed the noise as much as it can to achieve a better prediction result.

References

[1] Chen, P: Empirical and Theoretical Evidence of Economic Chaos, System Dynamics Revie*w*, 4 1988

[2] Dechert, W.D: Chaos Theory in Economics: Methods, Models and Evidence, Edward Elgar Publishing Limited, Cheltenham1996

[3] Grassberger, P. and Procaccia, I: *Physics Review Letters,* 50, 346, 1983

[4] Grassberger, P. and Procaccia, I: *Physics Review Letters,* 13, 34, 1984

[5] Peitgen, H., Jurgens, H., and Saupe, D: Chaos and Fractals, New Frontiers of Science, Springer-Verlag New York, Inc., New York, 1992.

[6] Peters, E.E: Chaos and Order in the Capital Market, John Wiley & Sons, Inc., New York, 1991.

[7] Urbach, R.M.A: Footprints of Chaos in the Markets, Financial Times Prentice Hall, London, 2000

[8] Wang, J. and Weston R: Perspectives On The Exploration Of Chaotic Structures In Stock Markets, Proceedings of the Conference on Advanced Investment Technology 1999, Bond University, Gold Coast, Australia December 19-21, 1999 (p.306) 1999.