Zagreb International Review of Economics & Business, Vol. 1, No. 1, pp. 61-82, 1998

ISSN 1331-5609

Testing for Nonlinearity and Deterministic Chaos in Monthly Japanese Stock Market Data

Teo Jašić*
University of California at Berkeley

Abstract: It has been widely recognised that the randomness of a stock market may actually be an indicator of an underlying strange attractor which has a fractal structure and supports chaotic motion. The application of non-linear methods to such financial data may indicate the presence of nonlinearities and low-dimensional chaos. These methods include rescaled range (R/S) analysis, correlation dimension calculation and estimation of Lyapunov exponents. This study presents a preliminary analysis of these tests when applied to the monthly TOPIX data of the Tokyo Stock Exchange. Although there are a number of limitations for applied non-linear methods such as the presence of noise and limited data size, the results indicate the presence of nonlinearities and the long memory effect in the observed data set. In order to complement these methods, neural networks are used for non-linear modelling and the prediction of the TOPIX data. The results can serve as an additional evidence of a deterministic system by giving accuracy estimates for short-term prediction.

JEL Classification: C 8

***Key words**:* non-linear analysis, chaos, neural networks, prediction

Introduction

The finding that a system which appears to behave randomly may in fact be a low-dimensional deterministic non-linear system has motivated the application of non-linear methods to financial data (Hinich *et al*, 1985; Scheinkman *et al*, 1989). Consequently, recent research of the stock market data has used techniques developed in physics to analyse nonlinear systems in order to discover non-periodic cycles governed by an underlying strange attractor. An analysis using these techniques may

* Teo Jašić is a Ph.D. candidate at the Department of Industrial Engineering and Operations Research, University of California at Berkeley, CA.

reflect the presence of important nonlinearities in the business fluctuations of the economic system. In practice, a deterministic system is often termed 'chaotic' if it shows sensitive dependence on initial conditions and has a relatively small number of degrees of freedom.

Studies in this area include the work of Brock, Hsieh and LeBaron (1991) who found non-linear dependence in the CSRP (Centre for Research in Security Prices) value weighted US stock returns index. Peters (1991) provided additional evidence that the S & P 500 index has an underlying low dimensional chaotic attractor along with similar findings for the data comprising UK, German, and Japanese market indices. Japan is cited as an interesting case as several studies have found an important non-linear structure in the Japanese market (Frank *et al*, 1989). The common methodology used in these studies is the BDS-test (Brock, Dechert and Scheinkman, 1988). The BDS-test is based on the correlation dimension statistic which contrasts the null hypothesis of independence and identical distribution of time series against a general alternative of dependence, either deterministic or stochastic.

In this study, several algorithms for the analysis of non-linear and chaotic time series are applied to observed TOPIX (Tokyo Security Exchange Stock Prices Index) data. The *rescaled range (R/S)* analysis is used to detect long-memory effect in the TOPIX time series over certain time period. In addition, the *correlation dimension*, which provides an estimate of the fractal dimension of the system, is being calculated by the Grassberger and Procaccia method (1983a). The lower bound for Kolmogorov entropy (K_2, Renyi entropy of second order), which is a useful indicator of system behaviour, can be calculated as the by-product of correlation dimension calculation. As a final test, the largest *Lyapunov exponent* is calculated to provide an additional evidence of the possible existence of a strange attractor in the Japanese Stock Market.

The aim of these algorithms is to calculate geometric and dynamical invariants of an underlying strange attractor, such as correlation dimension and Lyapunov exponents. The largest Lyapunov exponent is an indicator of how far into the future reliable predictions can be made, and the correlation dimension is an indicator of the complexity of a possible predictive model (Casdagli, 1989). It should be noted from the outset that the application of these algorithms is plagued with several problems, of which the most prominent are the limited size of data sets and the presence of noise in the data. Moreover, a number of researchers emphasise the necessity of exact and precise statistical analysis when using these algorithms which will make findings unambiguous (Jensen, 1993).

In order to complement these non-linear methods and to test for *determinism*, a neural network prediction model is also constructed. The observations of the prediction error by using neural networks to approximate given data set may provide insights about the deterministic properties of the observed data set.

Algorithms for non-linear Analysis

R/S analysis

The rescaled range analysis (R/S analysis) (Hurst,1951) is able to distinguish a random series from a non-random series, irrespective of the distribution of the underlying series (Gaussian or non-Gaussian). It is a robust statistics for measuring the amount of noise in a system and can be used to determine the average length of nonperiodic cycles. R captures the maximum and minimum cumulative deviations of the observations x_t of the time series from its mean (μ), and it is a function of time (the number of observations N)

$$R_N = \max_{1<t<N}\left[x_{t,N}\right] - \min_{1<t<N}\left[x_{t,N}\right] \tag{1}$$

where

$$x_{t,N} = \sum_{t=1}^{N}(x_t - \mu)\,,\ t=1,\ldots,N \tag{2}$$

The R/S ratio of R and the standard deviation S of the original time series can be estimated by the following empirical law (Mandelbrot,1972): $R/S = N^H$ when observed for various N values. For some value of N, the H exponent can be calculated by:

$$H = log(R/S)/log(N), \qquad 0 < H < 1. \tag{3}$$

and the estimate of H can be found by calculating the slope of the *log/log* graph of R/S against N using regression. The R/S used in the regression between *log(R/S)* and *log(N)* for various N, is the average R/S value for each N. By regressing over the range of different values of N, the highest value obtained indicates the average length of the nonperiodic cycle (mean orbital period) of the observed system.

The Hurst exponent H describes the likelihood that two consecutive events are likely to occur. There are three distinct classifications for the Hurst exponent: (a) $H = 0.50$, (b) $0 \le H < 0.50$, and (c) $0.50 < H < 1.00$. The type of series described by $H = 0$ is random, consisting of uncorrelated events. However, value $H = 0.50$ cannot be used as an evidence of a Gaussian random walk, it only proves that there is no evidence of long memory effect. A value of H different from 0.50 denotes the observations that are not independent. When $0<H<0.5$, the system is an anti-persistent or ergodic series with frequent reversals and a high volatility. For the third case ($0.5<H<1.0$), H describes a persistent or trend-reinforcing series which is characterised by long memory effects.

Correlation Dimension and Kolmogorov Entropy

The fractal dimension of given time series can be calculated by using the method of state-space reconstruction with time delay co-ordinates (Packard *et al.*, 1980). To extract information from a time series, it is necessary to geometrically reconstruct the attractor on which the trajectory producing the data is assumed to lie. This is achieved by constructing (embedding) m-dimensional vectors from a single co-ordinate (any single output of a dissipative dynamical system) and using time delays (lags) of the same variable. For the noise free data, it means that the topological properties of the attractor of the original system are the same as those for the embedded system (Takens, 1981). The details of the attractor may change, but the topological characteristics or the geometrical form remain intact. This is an important property as it implies that all systems are observable, regardless of whether the proper states needed to form an attractor are known or measurable.

The reconstructed phase space has all the characteristics of the real, original phase space, given that the time delay τ (in units of orbital periods) and embedding dimension m are properly specified. The relationship $m_\tau \approx 1$ is often used in reconstructions for the calculation of correlation dimension and Lyapunov exponent (Wolf *et al.*, 1985). If the time lag is represented in the units of observed time periods l, then $Q = ml$ and $\tau = l/Q$ where Q is the mean orbital period which can be estimated from the R/S analysis. It is the length of time until observations become uncorrelated.

The fractal dimension estimate provides important information about the underlying attractor. Grassberger and Procaccia (1983a) estimated the fractal dimension as the correlation dimension, D, which measures how densely the attractor fills its phase space by finding the probability that any one point will be a certain distance, r, from another point.

The correlation integral, $C_m(r)$, for a time series $\{X_t: t = 1, \dots N\}$ of m-dimensional vectors is an estimator of the probability that two vectors of the time series of length m are within a distance r of each other:

$$C_m(r) = \lim_{N\to\infty} \frac{2}{N_m(N-1)} \sum_{t<s} I_r(X_t^m, X_s^m), \tag{4}$$

where $N_m = N-(m-1)$, m is embedding dimension, and $I_r(x, y)$ is an indicator function which equals 1 if $\|x-y\| < r$, and 0, otherwise $\|x-y\|$ is the sup-norm (i.e., the L^∞ norm) which serves as a distance measure.

The correlation dimension for embedding dimension m is defined as:

$$D_m = \lim_{\substack{r\to 0 \\ N\to\infty}} \log C_m(r,N) / \log r \tag{5}$$

and the correlation dimension itself is

$$D = \lim_{m \to \infty} D_m \tag{6}$$

The correlation dimension can be used to differentiate between an apparently random system that is a low-dimensional deterministic chaos and a high dimensional or stochastic system. If chaos is present, i.e. the original time series is generated by an attractor of finite dimension, then the dimension estimate D_m will stabilise at some value for increasingly larger values of the embedding dimension m. If this stabilisation does not occur, the system is considered 'high-dimensional' or stochastic. For purely stochastic system the correlation dimension does not converge and is equal to m.

The estimate $C_m(r,N)$ of the correlation integral has an asymptotic behaviour (for m sufficiently large and r small) of the following form (Grassberger and Procaccia, 1983b):

$$C_m(r) = \lim_{N \to \infty} C_m(r,N) \sim r^D \exp(-mrK_2) \tag{7}$$

where K_2 is a lower bound for the Kolmogorov entropy K, which has the following properties: $K_2 \leq K$, $K_2 \geq 0$, K_2 is infinite for random systems, and $K_2 \neq 0$ for chaotic systems. $K_2 > 0$ is a sufficient condition for chaos and is found to be numerically close to Kolmogorov entropy for many dynamical systems. Thus, if $\log C_m(r,N)$ is plotted against $log\ r$ for increasing values of m, and assuming that the correlation dimension D exists, a series of approximately parallel lines with common slope D will be obtained, and the displacements between these lines will be multiples of the factor $\exp(-mrK_2)$. In particular,

$$\log C_m(r,N) - \log C_{m+1}(r,N) = rK_2 \tag{8}$$

and K_2 can be estimated from this formula, as

$$K_2 \sim \lim_{\substack{m \to \infty \\ r \to 0}} K_{2,m}(r) \tag{9}$$

The Spectrum of Lyapunov Exponents

Chaotic attractors are characterised by the sensitive dependence on initial conditions. The Lyapunov exponents are related to the expanding or contracting nature of different directions in phase space. A *positive* exponent indicates a 'direction' in which the system exhibits the repeated stretching and folding that decorrelates nearby states on the attractor. Such behaviour is chaotic because the long-term evolution of an initial condition that is specified with any uncertainty cannot be predicted. The *largest* Lyapunov exponent represents the divergence of points in phase space, or the sensitive dependence on the conditions represented by each point. For a chaotic attractor, it is greater than zero. The *i-th* one dimensional Lyapunov exponent (L_i) is defined in terms of the length of the stretched principal axis $r_i(t)$ (Wolf *et al*,1985):

$$L_i = \lim_{t \to \infty} \frac{1}{t} \log_2 \frac{r_i(t)}{r_i(0)}. \tag{10}$$

The importance of Lyapunov exponents lies in the fact that they quantify an attractor's dynamics in information theoretic terms. The exponents measure the rate at which system (dynamics) processes create or destroy information (Shaw, 1981). Therefore, the exponents are usually expressed in bits of information/s or bits/orbit for a continuous system.

In this study, the method of Wolf *et al.* (1985) for calculating the largest Lyapunov exponent L_1 from experimental data is used. It measures the divergence of nearby points in the reconstructed phase space, and indicates how the rate of divergence scales over fixed intervals of time. In reality, we deal with a limited amount of nonstationary and noisy data, so that the embedding dimension m, the time lag τ, and the maximum and minimum allowable divergence distance must be chosen appropriately. The relationship $m_\tau = 1$ is used in reconstructions, where m is the embedding dimension and $_\tau$ is the delay in units of orbital periods.

Results of the non-linear Analysis for the TOPIX Data

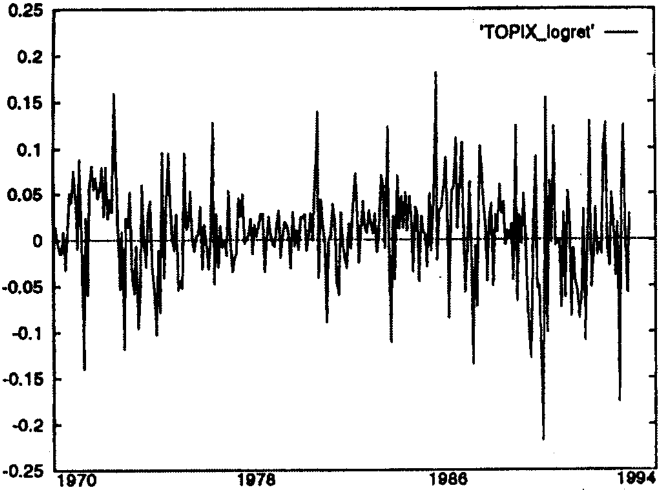
Data Summary

Data used in this study consist of the monthly TOPIX time series covering the period between March 1970 and March 1994. The whole set contains 289 monthly observations.

For market analysis, percentage changes or logarithmic returns defined as: $Si=log(P_{i+1}/P_i)$ are commonly used. For R/S analysis, logarithmic returns are more appropriate than the more commonly used percentage change in prices because R/S analysis is the cumulative deviation from the average and logarithmic returns sum to cumulative returns which is not the case for percentage changes. Figure 1 shows the graph of logarithmic monthly returns of the TOPIX.

It should be pointed out that there is a severe data limitation in finance with respect to data size and the sampling frequency. As one extends a data-set back in time, non-stationarity becomes increasingly more likely. Thus, the requirements of long sampling intervals (to avoid micromarket structure dependencies) and short histories (to avoid micromarket structure dependencies) and short histories (to avoid non-stationarity) impose severe data limitations in finance (Brock *et al*, 1992). In the analysis of nonlinear and chaotic systems, it is advantageous to have data covering longer time periods rather than large number of data points in a shorter period because time series may not be invariant with respect to time (i.e., 'time arrow' effect in the nonlinear systems).

Figure 1: Logarithmic monthly returns of the TOPIX



In the case of the calculation of correlation dimension of the time series, returns are not appropriate for non-linear dynamic analysis because by using returns the data is prewhitened and the serial dependence between observations may be eliminated (Theiler *et al*, -1992). It may be the case that this - serial dependence can reveal non-linear dependent structure through the noise. An additional problem for economic time series involves de-trending prices for economic and inflationary growth. If data for economic and inflationary growth are not available, the simple procedure which will log the price and subtract appropriate value depending on the time horizon may be used. For the case of the TOPIX data for which the inflationary data were not available, we use the following formula:

$$S_i = \log_e(P_i) - (a * index + Constant) \qquad (11)$$

where P_i represents the original time series, S_i denotes detrended price series, and *index* is the ordinal index of time series data points, *a* is a constant, and for *constant* the mean of the series $log_e(P_i)$-a*$index$ can be used. The resulting loglinear detrend time series of the Topics is shown in Figure 2.

Figure 2: Time 5eries of loglinear detrended TOPIX: March 1970 - March 1994

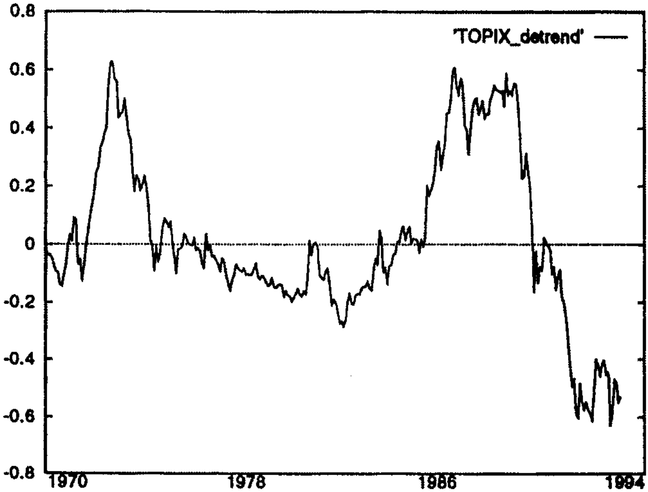


Table l: Regression between *log(R/S)* and *log(N)*

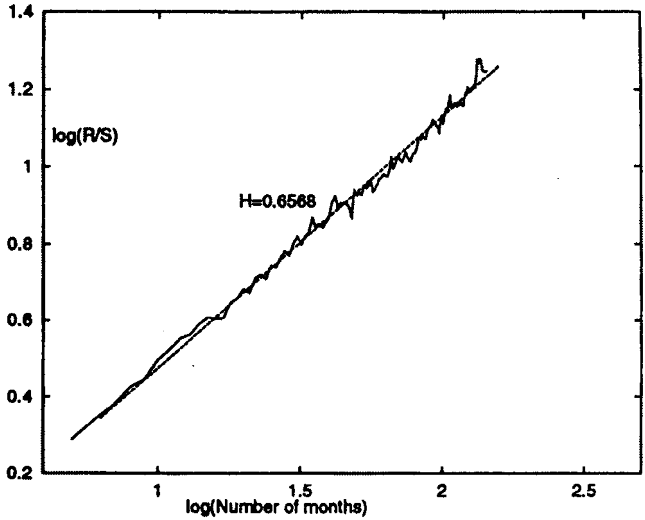
R^2	0.99122
Adj. R^2	0.99116
St. Error	0.02093
F Test	15475.52
Sig. F.	0.0000
X coef. β	**0.656899**
Constant	-0.182272

Results of non-linear Analysis

Results of R/S Analysis

The results of applying the R/S analysis to the log-returns of the TOPIX data are shown in Table l. The *H* value is found to be 0.6568 (β coefficient of X, slope estimate). It is calculated by using regression between the *log(N)* and *log(R/S)* over the whole range of N values. The H value of 0.6568 indicates an evidence of non-linear dependence in the series although a high level of noise seems to be present.

Figure 3: R/S Analysis of Japanese TOPIX index: Monthly Returns, March 1970 - March 1994



By running regression over the range of data which increases with the number of data points corresponding to 2, 3,...,10 years, the largest slope estimate is obtained for the period of four years, although the values for neighbouring periods (3 or 5) are not significantly lower. Consequently, for the observed data set, the TOPIX returns have a mean orbital period or mean memory period of approximately 4 to 5 years.

Figure 3 shows the plots of *log N* against *log(R/S)* for the Japanese market. The point on x-axis *(log N)* for which the *R/S* observations begin to fluctuate and become erratic is close to the average cycle length of the system. It seems that the Japanese market has a four to five year cycle on the account of the data available, although it cannot be readily seen from the *log/log* plot. The cycle length can be associated with the economy cycle of the country but the nature of this relationship is not easily determined.

The validity of the H estimate can be tested by randomly interchanging the order of data points in original time series and calculating the H exponent for a new series. For the long memory effect, the order of data is important so that a new series should have lower H estimate, although the frequency distribution of the observations remains unchanged. After shuffling original time series to ten new series, the average obtained H estimate was much lower (0.51).

Calculation of Correlation Dimension

Using the results of the *R/S* analysis for the definition of the appropriate time lag, the correlation integrals *C(r)* are calculated for embedding dimensions from 3 to 15. The initial distance *r* was chosen to be 10 per cent of the amplitude range of the original time series. Figure 4 shows the *log/log* plots of correlation integrals for embedding dimensions of 3 to 15. The regression is run for each dimension over the linear regions of the *log/log* plots. The correlation dimension should eventually converge to its true value as the embedding dimension *m* is increased. Figure 5 shows the results of the regression between *log C(r)* and *log r*, which are the estimated correlation dimensions D_m for each dimension *m* (represented as points on the lower curve). It can be noted that correlation dimension values (D_m) converge to a value of 2.355 for increasing values of dimension *m*, so $D \approx 2.355$.

In order to validate the results of correlation dimension estimation, a new time series is generated from the original data by choosing data at random with replacement. For the reordered data the slopes of correlation integrals will increase with *m* indefinitely since the structure of the data is lost in permutation (Isham, 1993). If the original time series is random then the order of the time series will not change the value of the correlation dimension. The correlation dimension statistic for the 'randomised' loglinear de-trended TOPIX data can be distinguished from the same statistic for original data as shown in Figure 5. Observation of the actual values of the statistic for original and 'randomised' data reveals that difference can be greater than 10 per cent. It means that there is evidence of a non-linear structure in the original time series. This procedure should be viewed as refuting some hypothesis about the time series, rather than proving the evidence of chaotic system.

From the correlation dimension calculations, the lower bound K_2 for Kolmogorov entropy as given in Formulas 8 and 9 can also be estimated. Figure 6 shows the K_2 values for the observed data for the first three values of *r* taken in the linear region of slopes (see Figure 4). The extrapolated value seems to be lower than the value of the Lyapunov exponent (for the linear region of the correlation dimension integrals, the value of K_2 is in the range of 0.015 to 0.03). The problem with this method is that the results are sensitive on choices of *r* and may diverge for higher embedding dimensions and short time series.

It is important to note that a method of correlation dimension calculation by fitting a straight line to the plot of *log Cm (r, N)* against *log r* has some deficiencies such as the wandering intercept problem (Cutler, 1991) and the effects of the observation noise at small distances of *r*. There is in principle a trade-off between taking *r* small enough to avoid non-linear effects and taking *r* sufficiently large to reduce statistical errors due to lack of data (Isham, 1993).

Figure 4: Correlation Integrals for loglinear detrended TOPIX data

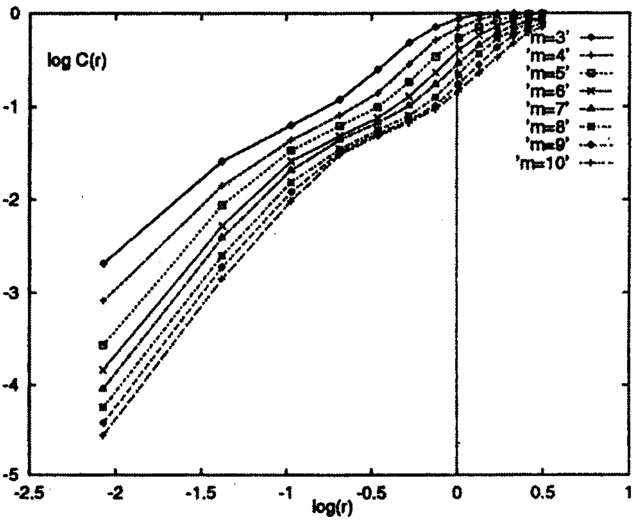


Figure 5: Estimate of the correlation dimension for loglinear detrended TOPIX data and comparison with randomised TOPIX data.

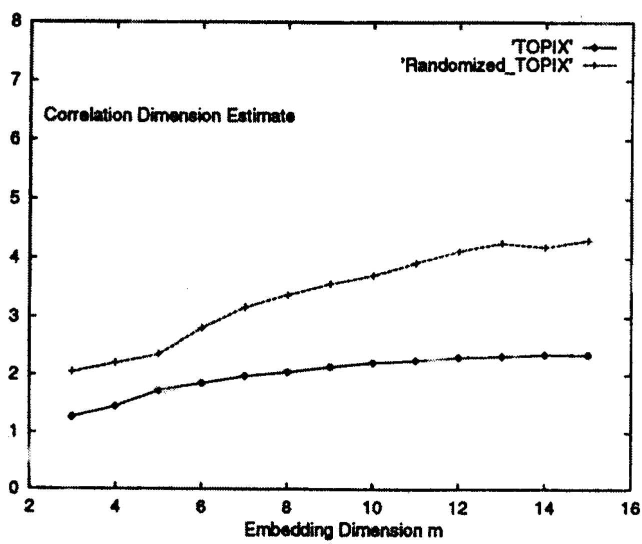
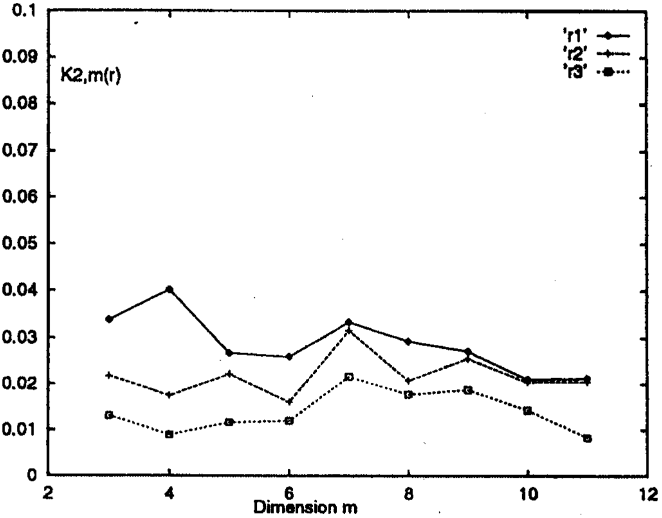


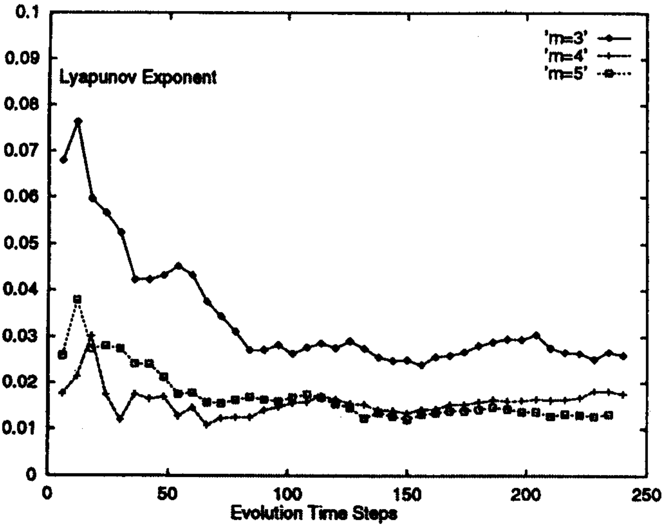
Figure 6: Extrapolated values of $K_{2,m}\ (r)$ for loglinear detrended TOPIX data



Calculation of the Largest Lyapunov Exponent

The method for the calculation of the largest Lyapunov exponent is applied to a 3,

Figure 7: Convergence of the largest Lyapunov exponent



4, and 5-dimension reconstructed phase space because the estimated value of the correlation dimension is $D = 2.355$. Embedding the attractor in a higher dimensional space is desirable and results in stretching the attractor and reducing systematic errors due to a non-uniform coverage of the attractor (Isham, 1993). Evolution time steps T of 4 and 6 months were tried for the experimental data set in order to avoid folds, and the maximum divergence between two points was chosen to be 10 per cent of the amplitude range of the time series.

Figure 7 shows the stable convergence of the estimated values of the largest Lyapunov exponent given the evolution time step of 6 months to 0.026698 for dimension $m = 3$, and to 0.017057 and 0.013508 for dimension 4 and 5 respectively. If the evolution time step is taken to be 4 months, the Lyapunov exponent estimates are slightly higher for the observed dimensions (3, 4, and 5).

Summary of Results for non-linear Methods

The purpose of the applied algorithms was to search for the presence of important nonlinearities in the TOPIX monthly data. The results should be viewed as an exploratory data analysis. The results of the R/S analysis indicate that the long memory effect is present, with an average cycle of 4 to 5 years. Similar result is also presented in the calculation of the largest Lyapunov exponent with a value of approximately 0.0267 bit/month. Therefore, the whole information is expected to be lost in $1/0.0267 \approx 38$ months (for an embedding dimension 3) or $1/0.017057 \approx 58$ months in the case of the embedding dimension of 4. Correlation dimension estimate of 2.355 shows that definite temporal correlations in the data do exist and provides an indication of the complexity of the system.

The Neural Network Prediction Model Approach

The algorithms applied in previous sections help to calculate geometric and dynamical invariants of an underlying strange attractor. However, the data requirements for these algorithms are often prohibitive and the calculated invariants are of limited practical use. Another approach to determine some of the properties of the time series is to construct a prediction model directly from time series data whereas the prediction will be treated as an 'inverse problem' in dynamical systems (Casdagli, 1989, 1991; Farmer and Sidorowich, 1987). The essence of the inverse problem is to construct a non-linear map, which is able to capture the asymptotic behaviour of the observed time series. Such a map, given initial assumptions and satisfactory performance criteria, can serve as a candidate for a prediction model. The ability to forecast accurately with such a model is the test of whether or not low-dimensional chaos is present. This method can be used to complement previously described algorithms.

Embedding with State Space Reconstruction

It is assumed that the time series observation x_t is a scalar variable sampled at discrete intervals of time $t = n\tau,\ 1 \le n < \infty$ and that the underlying dynamics is that of a strange attractor lying on a (possibly) low-dimensional invariant manifold of a dynamical system. Using state space reconstruction, a smooth map $f: R^m \to R$ can be constructed for observed values of the finite time series $x_t, 1 \le n \le N$:

$$f(x((n+m-1)\tau, \dots, x(n\tau)) = x((n+m)\tau). \tag{12}$$

A value m in the formula is referred to as an embedding dimension and it is possible to find the minimal embedding dimension m^* for which the formula holds. For a deterministic system, it is sufficient to take $m > 2D$ for equation (12) to hold. The results of Sauer *et al* (1991) also show that if $D < m < 2D$, equation (12) can also hold where D is the dimension of the reconstructed attractor. On the other hand, if the time, series is generated by a stationary stochastic process then a noise term will appear on the right-hand side of equation (12) (for all values of m) and the delay vectors are expected to reconstruct a set of dimension m.

Evaluation of the Prediction Errors

The prediction error can be evaluated either as the normalised mean squared error (NMSE) or the normalised root mean squared error (NRMSE) of the testing (prediction) set P described as follows:

$$E_m(P) = \frac{1}{\sigma_P^2}\frac{1}{N}\sum_{n\in P}(x_n - \hat{x}_n)^2 = \frac{1}{\sigma_P^2}\frac{1}{N}\sum_{n\in P}(x((n+1)\tau) - f_n(x(n\tau), \dots, x((n-m+1)\tau)))^2 \tag{13}$$

where σ_P^2 is the estimated variance of the time series for a testing set P, N is the number of data points in a set P, x_n is the actual value and $\hat{x}_n$ is the predicted value for certain n in set P. E_m is the normalised mean squared error and $E_m^{1/2}$ is the normalised root mean squared error.

Although the discrimination between deterministic or stochastic models is often not clear-cut, the method based on prediction can be used to complement algorithms described in the previous sections. The procedure consists of evaluating the prediction error for increasing values of m until an acceptably small value is found for E_m. In the case of chaotic time series, it is expected that the prediction errors E_m will suddenly decrease to a value close to zero as m is increased to the correct minimal embedding dimension m^* and remain close to zero for values of m above m^*. If the time series is random, no such decrease should be observed (Casdagli, 1989). An assumption here is that m is large enough for the map to be an embedding so, in the deterministic case, improved forecasting will occur as m increases until the appropriate value of m^* is reached and then no further improvement will result.

In addition, the prediction errors can be observed for values of m and N (number of observations in the testing set), as the function of the number of iterations in a multi step prediction. For a chaotic system, E_m should increase exponentially with the increasing number of iterations, at least for a while (Casdagli, 1992).

A limitation for the prediction method in general is that the sampling rate r must be kept reasonably small, so that the approximated function does not have wildly varying slopes (Casdagli, 1989). In this study, time delay is 1.

Prediction with Neural Networks for Analysis

The numerical techniques appropriate for this problem must be able to interpolate or approximate unknown functions from scattered data points. The common used techniques are piecewise linear approximators, radial basis functions and neural networks. With using data of limited size, smoothing methods such as thin plate splines or neural networks are found to be most appropriate (Casdagli, 1989).

For this experiment, feed-forward single hidden layer neural networks are used, consisting of input units, hidden units, and one output unit: The input units correspond to time-delay vectors of a certain embedding dimension m, and the output unit represents predictions for the actual time series observations. A neural network with a sigmoid activation function is performing an operation similar to a multidimensional spleen approximation (Lapedes, Farber and McCelland, 1987), such that any arbitrary nonlinearity is sufficient to approximate any function. The nonlinearities are located in the non-linear transfer function of the hidden units, which is *tanh*.

Neural network units perform a weighted sum of the output of the preceding layer and in the addition the hidden units carry out a transformation by a sigmoidal transfer function g. If the input to the network are observations $x_{i-m+1},\dots,x_i$ then the output h_k of the k units in the hidden layer and the output o are given by:

$$h_k = g(\sum_{j=1}^{m} w_{kj} x_{i-j+1} + \theta_k) \tag{14}$$

$$o = \sum_k w_k h_k + \vartheta_o \tag{15}$$

where w_{kj}, θ_k, w_k, and ϑ_o are parameters of the system. In a training phase these parameters are adjusted in order to produce the desired output using the back propagation algorithm (Rumelhart, 1986). The mean squared error between the output (with a chosen number of hidden units *k* and function *g*) and the observed value is to be minimised.

Results of the Neural Network Prediction Model Approach for the TOPIX Data

Training and Testing Data Sets

Training and testing data for neural network are derived from the 289 samples such that data from the earlier period (1970/1989 bottom data) are used for training (200 samples) while the rest of data sets (1989-1994) are used for testing to observe the prediction error. Input vectors are adjusted to zero mean and unit variance for the neural network inputs in the interval (-1,1) (*tanh* activation function). In contrast to artificially generated data, the economic time series are limited in size so that conclusions can be made only for the allocated testing set.

For our experiments, the values of m from 3 to 10 are used and the prediction error is observed for both *single-step* prediction and *multi step prediction method.* The networks are trained with different learning rates in order to achieve optimal convergence. The results presented here are obtained using a learning rate of 0.05 and no momentum. Different neural network architectures which vary the number of hidden nodes are also tried out. The networks are chosen with the number of hidden units being less than the number of inputs.

Results of the Neural Network Prediction Model for Single-Step and Multi-Step Prediction

To test the prediction errors of trained neural network, two types of predictions are used: single-step prediction and multi-step prediction.

The term *single step prediction* refers to the use of the actual values of the observed time series for all input units. The normalised mean squared error is used to evaluate the single-step prediction performance. Given in the form (13), it is independent of the dynamic range of the data and of the length of time series. In practice, this method will yield less smooth prediction values but it can help evaluate the adaptability and robustness of the prediction model.

In *multi step prediction*, the set of predicted rather than actual values is used as the input of delayed input vectors to predict the value of a target in the next period. The network predicts the output variable one step ahead of time, but uses the predicted rather than the actual value for the current prediction. The prediction error in this case will depend on the number of iterations.

The resulting prediction errors for single-step prediction (on a testing set starting from index 200) are shown in Table 2. For $m = 6$, the lowest estimate of the prediction error is obtained (0.123201) although the error for $m = 4$ does not differ significantly. For $m \geq 7$, the prediction errors start increasing. Figure 8 shows the predicted and actual values over the whole data set for $m = 6$ using neural network with 6 inputs and 4 hidden units (6x4x1 network). The prediction errors are observed for the values of m such that $D < m \leq 2D$ and $m > 2D$ where D is the dimension of the reconstructed attractor (2.355 in this case). For a (stationary) stochastic process, the embedding will include a noise term and the error will increase with increasing values of m. In our case, the prediction errors for $m = 4$ and $m = 6$ decrease when compared to prediction errors for $m = 3$ and $m = 5$ respectively. This behaviour is not likely for stochastic systems and is consistent with the observations of Casdagli (1989) for chaotic systems.

For multi-step prediction, the prediction errors for networks corresponding to $m =$ 3, 4, 5, and 6 increase first exponentially for the increased number of iterations (up to 10) and then at a slower rate (nearly linear). For $m \geq 7$, the prediction errors increase at a very slow rate in the beginning (for the first 10 iterations) and then increase at a faster rate (almost exponentially). Figure 9 shows the values of the normalised root mean squared error (NRMSE) after 30 iterations for different values of m. In this case, the nearly exponential rate with which prediction errors increase for $3 \leq m \leq 6$ and the increased number of iterations may serve as an indication of a deterministic behaviour.

Table 2: Prediction Error (Testing) Results of Neural Networks (NN) for single step prediction for networks trained at a learning rate of 0.05

Model	m	NMSE
NN 3x2x1	3	0.146067
NN 4x2x1	4	0.126875
NN 5x3x1	5	0.139057
NN 6x4x1	6	0.123201
NN 7x4x1	7	0.136080
NN 8x4x1	8	0.137440
NN 9x5x1	9	0.156360
NN 10x6x1	10	0.168904

Figure 8: The normalised TOPIX data fitted with 6-4-1 network and the single step prediction of the network (Testing Set starts at the index point 200 on the x axis)

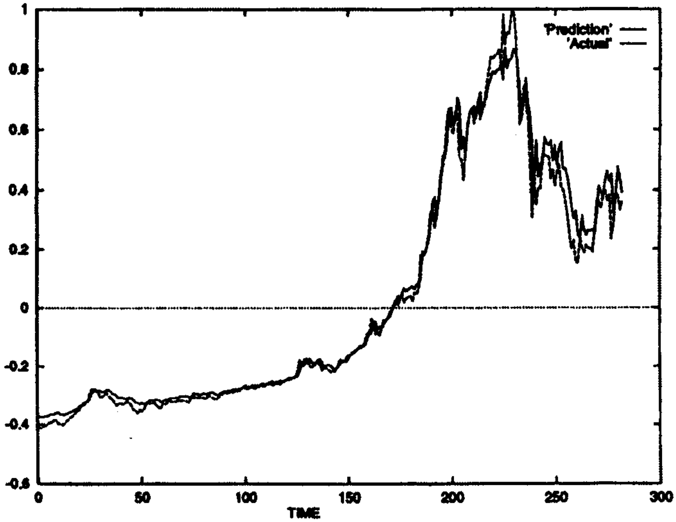
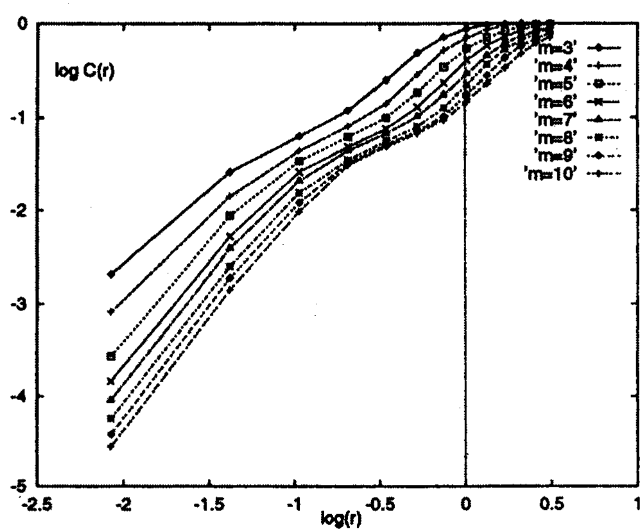


Figure 9: Normalised roots mean squared error after 30 iterations for different values of embedding dimension (input units to neural network) for multi-step prediction



According to these results, there is a certain indication of determinism when evaluating the prediction errors for single-step and multi-step prediction for the values of m in the observed range of 3 to 10.

Conclusion

In this study, non-linear methods for testing the non-linear structure of time series have been applied to the TOPIX monthly data. The correlation dimension estimate shows convergence and indicates temporal correlations in the data. The R/S analysis provides evidence that the observed system has a long-memory effect. Finally, the calculation of the largest Lyapunov exponent corresponds to the result of R/S analysis and may indicate sensitivity on the initial conditions in the system. Alternatively, in order to complement the above algorithms in detecting the deterministic properties of the TOPIX data, the prediction errors resulting from the neural networks approximation of the data set are evaluated. The application of the neural networks shows that the system exhibits certain deterministic properties for the allocated testing set. The unified approach of combining both the applied non-linear algorithms and neural network prediction model is necessary as the discrimination between deterministic or stochastic systems is not straightforward.

As applied to the TOPIX data, the above methods indicate a weak chaotic behaviour in the TOPIX monthly time series taking into account a limited size of the data set and the presence of noise. However, the results show that 'a certain level of deterministic behaviour of low dimensionality is present' in the TOPIX monthly time series.

Future research should adopt a more rigorous approach to testing the statistical significance of these applied methods for detecting chaotic behaviour in the data.

REFERENCES

Brock W. A., W.D. Dechert, J.A. Scheinkman, and B. LeBaron, (1988), *A Test for Independence Based on the Correlation Dimension,* (Madison, WI: University of Wisconsin Press).

Brock W.A., D.A. Hsieh and B. LeBaron (1991), *Nonlinear Dynamics, Chaos, and Instability, Statistical Theory and Economic Evidence* (Cambridge, MA: MIT Press).

Casdagli, M (1989), 'Nonlinear Prediction of Chaotic Time Series', *Physica*, 35:335-356.

Casdagli, M (1991), 'Nonlinear Forecasting, Chaos and Statistics', in L. Lam (Ed.), *Modelling Complex Phenomena* (Berlin: Springer-Verlag).

Casdagli, M (1992), 'Chaos and Deterministic versus Stochastic and Non-Linear Modelling', *Journal of Royal Statistical Society*, 54:303-328.

Cutler, C., (1991), 'Some Results on the Behaviour and Estimation of the Fractal Dimensions of Distributions on Attractors', *Journal of Statistical Physics*, 62:651-708.

Farmer, J.D., J.J. Sidorowich, (1987), 'Predicting Chaotic Time Series', *Physical Review Letters*, 59 (8):62-65.

Frank M.Z., R. Gencay, T. Stengos, (1988), 'International Chaos?', *European Economic Review, 32:1569-1584.*

Grassberger P., I. Procaccia, (1983a), 'Characterization of Strange Attractors', *Physical Review Letters*, 50:346-349.

Grassberger P., I. Procaccia, (1983b), 'Estimation of the Kolmogorov Entropy from a Chaotic Signal', *Physical Review Letters*, 28(4): 2591-2593.

Hinich M.J., D.M. Patterson, (1985), 'Evidence of Nonlinearity in Daily Stock Returns', *Journal of Business & Economic Statistics*, 3(1):69-77.

Hurst, H.E., (1951), 'Long-Term Storage of Reservoirs', *Transactions of the American Society of Civil Engineers*, 116.

Isham V., (1993), 'Statistical Aspects of Chaos: A Review', in O.E. Barndorff-Nielsen, J.L. Jensen, W.S. Kendall (Eds.), *Networks and Chaos - Statistical and Probabilistic Aspects* (London: Chapmann & Hall).

Jašić, T., (1995), 'Nonlinear Analysis of the Japanese Stock Market Data', *Proceedings of the 1995 International Symposium on Nonlinear Theory and Its Applications* - NOLTAÆ 95,10-14 December, Las Vegas, USA.

Jensen J.L., (1993), 'Chaotic Dynamical Systems with a View Towards Statistics: A Review', in O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (Eds.), *Networks and Chaos - Statistical and Probabilistic Aspects* (London: Chapmann & Hall).

Mandelbrot B., (1972), 'Statistical Methodology for Nonperiodic Cycles: From Covariance to R/S Analysis', *Annals of Economic Statistics and Measurement*, 1.

Packard N., J. Crutchfield, D. Farmer and R. Shaw, (1980), 'Geometry from a Time Series', *Physical Review Letters*, 45:712-716.

Peters E.E., (1991), *Chaos and Order in Capital Markets* (New York: John Wiley & Sons Inc).

Ripley, B.D., (1996), *Pattern Recognition and Neural Networks* (Cambridge: Cambridge University Press).

Rumelhart D. E and J.L. McClelland, (1986), *Parallel Distributed Processing*, Vol. 1 (Cambridge: MIT Press).

Sauer, T., J.A. Yorke and M. Casdagli, (1991), 'Embedology', *Journal of Statistical Physics*, 65(3/4):579-616.

Scheinkman J.A., B. LeBaron, (1989), 'Nonlinear Dynamics and Stock Returns', *Journal of Business*, 62(3):311-338.

Shaw, R., (1981), 'Strange Attractors, Chaotic Behaviour, and Information Flow', *Z. Naturforsch.*, 36 A:80.

Takens, F., (1981), 'Strange Attractors in Turbulence', in *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics no. 898 , (Berlin: Springer-Verlag) 366-381.

Theiler, J., B. Galdrikian, A. Longtin, S. Eubank, J.D. Farmer, (1992), 'Using Surrogate Data to Detect Nonlinearity in Time Series', in M. Casdagli and S. Eubank, (Eds.), *Nonlinear Modelling and Forecasting,* (Reading, MA: Addison-Wesley).

Wolf A., Swift J.B., Swinney H.L., Vastano J.A., (1985), 'Determining Lyapunov Exponents from a Time Series', *Physica*, 16:285-317.