Int. Computer Symposium, Dec. 15-17, 2004, Taipei, Taiwan.

# Detecting Long-range Power-law Correlations in Financial Time Series: A Case on Listed Companies of Taiwan Stock Market

Huei-Huang Chen Yi-Lin Huang
*Department of Computer Science and Engineering*
*Tatung University*
*40 Chungshan North Road, Section 3*
*Taipei, 104, Taiwan, R.O.C.*
hhchen@ttu.edu.tw g9106026@ms2.ttu.edu.tw

***Abstract**- In time series analysis, there have been many statistic models widely used; some models could estimate long memory. A new idea for analyzing time series is Detrended Fluctuation Analysis (DFA), which was originally developed for finding long-rage power-law correlations in DNA sequences. We apply DFA to Taiwan stock market for three categories of data: TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index), the group indices aggregated from individual stock indices, and individual stock indices. The results show that long memory exists in most listed companies of Taiwan stock market for the cases when $\alpha \neq 0.5$. However, the correlations detected from aggregated data series do not imply the correlation of original data series. Our findings are that the correlations detected from main index do not imply the same correlation of group indices and individual stock indices, but there are greater than half of group indices and individual stock indices following the same correlation with the main index.*

**Keywords:** Detrended Fluctuation Analysis, Time Series Analysis, Long Memory.

## 1. Introduction

In time series analysis, there have been many statistic models widely used; some models could estimate long memory and they are non-parametric, such as Hurst analysis [1]. If a series is detected as persistent correlation, its trend will be unchanged in the future and vice versa. Therefore, the existence of long memory in financial time series data would be a good reference for investment.

A new idea for analyzing time series is Detrended Fluctuation Analysis (DFA) [2], which was originally developed for finding long-rage power-law correlations in DNA sequences, and the advantages of DFA over conventional methods (e.g., spectral analysis and Hurst analysis) are that it permits the detection of intrinsic self-similarity embedded in a seemingly non-stationary time series, and also avoids the spurious detection of apparent self similarity, which may be an artifact of extrinsic trends [3]. Furthermore, this method has been applied to financial time series analysis in recent years [4][5][6].

Most previous researches on analyzing time series data of stock market using DFA focused on the main indices of each nation to detect long-range power-law correlation. Therefore, we applied DFA to analyze Taiwan stock market for three categories of data: TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index), the group indices aggregated from individual stock indices, and individual stock indices to see whether there exists some long-range power-law correlation in each of them. The results of the analysis in those three categories of indices are then compared to each other.

The rest of the paper is organized as follows: Section 2 presents literatures review in the field of long memory time series analysis; three long memory analysis methods are included. Section 3 details how we use the DFA method in our analysis. Section 4 describes the results, followed by the conclusions and future works in Section 5.

## 2. Literatures Review

Time series analysis has been used to various fields for many years. Methods to analyze time series could be applied to various domains, such as Hurst method that was originally developed for water resource researches and was also used in finance researches. Section 2.1 gives a brief introduction to time series analysis. Section 2.2 presents three long memory analysis methods.

### 2.1. Time Series Analysis

The analysis of experimental data that have been observed at different points in time leads to new and unique problems in statistical modeling and inference. The obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of many conventional statistical methods which depends on the assumption that these adjacent observations are independent and uniformly distributed. The systematic approach by which one goes about answering the mathematical and statistical questions posed by these time correlations is commonly referred to as time series analysis [7].

The measurement of some particular characteristic over a period of time constitutes a time series. It may be an hourly record of temperature at a given place or the annual rainfall at a meteorological station. It may be a quarterly record of gross national product; an electrocardiogram may compose several time series [8].

The theoretical correlation expresses the dependence of the time series observations on each other. This dependence can also be expressed by regression model that represents the present observation as the sum of two independent, uncorrelated, or "orthogonal" parts: one dependent on the preceding ones and the other an independent sequence [9]. A system is said to exhibit long-range correlations when some physical properties of the system at different positions (or times) are correlated and its correlation function decays much slower than exponential decay. This long-range correlation property can be a consequence of a diverging (infinite) correlation length or there is no characteristic length ("scale-free") for the correlation. In the latter case, the correlation is usually associated with power-law decay. The mechanisms for generating long-range correlations are not always obvious, but in most cases they can result from two different origins: (i) Long-range correlations are generated by some physical interaction of the system, the interaction is usually short-range but manifests itself to long-range correlation under special physical condition. One famous example is the magnetic system at critical point. (ii) Long-range correlations are simply a reflection of other scale-invariant properties of the system [10].

## 2.2. Long Memory Detecting Methods

Estimating long memory in time series is an important topic of time series analysis and there had been many approaches developed in this filed. In this section, three commonly used approaches of long memory analysis will be presented. We will give an introduction to rescaled range analysis and periodogram regression in lines of Weron's [11] work. Finally, we will discuss the detrended fluctuation analysis, which we will use in our experiments.

### 2.2.1 Rescaled Range Analysis

Rescaled range analysis, also called R/S or Hurst method was developed by Hurst [1] for the research of the variation of water capacity of river Nile in Egypt. His original work is related to the design of an ideal storage on river Nile. Using this method can finally obtain a parameter $H$ called Hurst exponent, which measures the intensity of long-range dependence in a time series.

The R/S method begins with dividing a time series of length $L$ into $d$ subseries of length $n$. Next for each subseries $m = 1, \mathbf{K}, d$ :

- Find the mean ( $E_m$ ) and standard deviation ( $S_m$ ).
- Normalize the data ( $Z_{i,m}$ ) by subtracting the sample mean $X_{i,m} = Z_{i,m} - E_m$ for $i = 1,...,n$ .
- Create a cumulative time series $Y_{i,m} = \sum_{j=1}^{i} X_{j,m}$ for $i = 1, \mathbf{K}, n$ .
- Find the range $R_m = \max\{Y_{1,m}, \mathbf{K}, Y_{n,m}\} - \min\{Y_{1,m}, \mathbf{K}, Y_{n,m}\}$ .
- Rescale the range $R_m / S_m$ .

Finally, calculate the mean value of the rescaled range for all subseries of $n$

$$(R/S)_n = \frac{1}{d}\sum_{m=1}^{d} R_m / S_m . \quad (1)$$

The R/S method asymptotically follows the relation

$$(R/S)_n \sim cn^H . \quad (2)$$

Thus the value of $H$ can be obtained by running a simple linear regression over a sample of increasing time horizons

$$\log(R/S)_n = \log c + H \log n . \quad (3)$$

The Hurst exponent ranges from 0 to 1 and it can be classified into 3 ranges:

- For $H = 1/2$ the time series is called uncorrelated random series.
- For $1/2 < H < 1$ the time series is probably persistent.
- For $0 < H < 1/2$ the time series is probably antipersistent.

### 2.2.2 Periodogram Regression

This method is a semi-parametric procedure to obtain an estimate of the fractional differencing parameter $d$ . This technique, proposed by Geweke and Porter-Hudak [12], was also called GPH method, based on observations of the slope of the spectral density function of a fractionally integrated series around the angular frequency $w = 0$ . Since they showed that the spectral density function of a general fractionally integrated model with differencing parameter $d$ is identical to that of a fractional Gaussian noise with Hurst exponent $H = d + 0.5$ , the GPH method can be used to estimate $H$.

The estimation procedure begins with calculating the periodogram, which is a sample analogue of the spectral density. For a vector of observations $\{x_1, \mathbf{K}, x_L\}$ the periodogram is defined as

$$I_L(w_k) = \frac{1}{L}\left|\sum_{t=1}^{L} x_t e^{-2pi(t-1)w_k}\right|^2 \quad (4)$$

where $w_k = k/L, k = 1, \mathbf{K}, [L/2]$ and $[x]$ denotes the largest integer less then or equal to $x$. Observe that $I_L$ is the squared value of the Fourier transform and if the observations vector is of appropriate length.

The next and final step is to run a simple linear regression

$$\log\{I_L(w_k)\} = a - d\log\{4\sin^2(w_k/2)\} + e_k \quad (5)$$

at low Fourier frequencies $w_k, k = 1, \mathbf{K}, K \le [L/2]$. The least squares estimate of the slope yields the differencing parameter $d$, hence $H = d + 0.5$. A major issue on the application of this method is the choice of $K$.

### 2.2.3 Detrended Fluctuation Analysis

C.-K. Peng et al. [2] developed the DFA method for their researches on DNA sequences. This method is independent of investigator input and permits the detection of long-range correlations embedded in a series, and also avoids the spurious detection of apparent long-range correlations that are an artifact of trends.

The DFA method comprises the following steps:

- Divide the entire sequence of length $N$ into $N/l$ non-overlapping boxes, each containing $l$ data, and define the "local trend" in each box (proportional to the compositional bias in the box) to be the ordinate of a linear least-squares for the DNA walk displacement in that box.
- Define the "detrended walk", as the difference between the original walk $y(n)$ and the local trend $y_l(n)$. Calculate the variance about the detrended walk for each box, and calculate the average of these variances over all the boxes of size $l$, denoted $F_d^2(l)$.

The function $F_d^2(l)$ is defined as following:

$$F_d^2(l) = \frac{1}{l}\sum_{n=kl+1}^{(k+1)l}(y(n) - y_l(n))^2, k = 0,1,2,\mathbf{K},(\frac{N}{l} - 1). \quad (6)$$

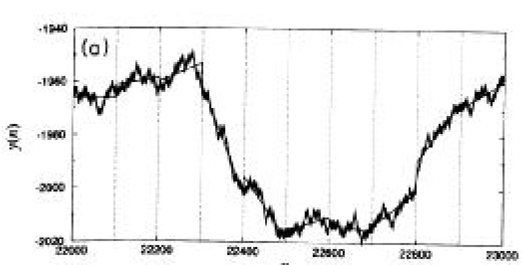


**Figure 1. DFA with box size *l* = 100**

To illustrate the DFA method, they use a 1000-nucleotide subsequence of the DNA sequences. Figure 1 shows the local trends when this subsequence is partitioned into boxes of size $l = 100$.

Figure 2 shows the local trends when the subsequence is partitioned into boxes of size $l = 200$. It is apparent by visual inspection that the variance increases with the box size. The dependence of variance on box size gives rise to the scaling properties of the fluctuations.

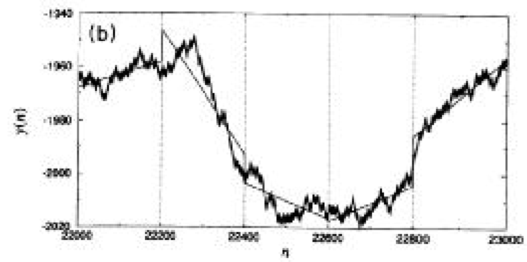


**Figure 2. DFA with box size *l* = 200**

If only short-range correlations (or no correlations) exist in the nucleotide sequence, then the detrended DNA walk must have the statistical properties of a random walk; the behavior is expected to be a power law $F_d(l) \sim l^a$ (7)

with an exponent $\alpha = 1/2$. An exponent $a \ne 1/2$ in a certain range of $l$ values implies the existence of long-range correlations in that time interval, i.e. if values of $a$ are less than $1/2$, the evolution of the events will follow an antipersistent power-law while values of $a$ is greater $1/2$ indicate persistence in the event sequence.

## 3. Research Methodology

### 3.1 Problem Description

Most previous researches on stock market using DFA focused on the main indices of each nation to detect long-range power-law correlation. They all proved that DFA could be used to estimate long memory in stock markets. However, detecting long-range correlations in main index is just useful for long-term investment in future market. If long-term investors want to invest individual companies, detecting long-range correlations in individual stock indices is also needed.

Moreover, we know that the main index and group indices are aggregated from individual stock indices. If the main index is analyzed as persistent correlation, what correlations of group indices and individual stock indices will be? Will the three categories of indices follow the same correlation or there has no relation between them? Therefore, we want to use DFA to analyzed Taiwan stock market with the three categories of data: TAIEX (Taiwan Stock Exchange

Capitalization Weighted Stock Index), the group indices, and individual stock indices.

### 3.2 Method Selection

In section 2, we presented three long memory analysis methods: R/S method, GPH method, and DFA method. We surveyed the literatures about the three methods and findings are all presented that DFA method is better than others.

In Weron's [11] work, he had tested these three methods on samples drawn from Gaussian white noise. The DFA method is more accurate on estimating the scaling exponent. Another similar work [13] compared the performance of the DFA method with Hurst method and show that the DFA is a superior method to quantify the correlation of noisy signals. We know that many non-economic factors (noisy signal) will influence the stock market in short-term period. So, the DFA can work better then Hurst method, because DFA can avoid spurious detection.

Moreover, the GPH method assumes that the data should be seasonal time series but our research focuses on daily closing price of listed companies. Therefore the GPH method may not work well in our research for its' assumption. So, We chose DFA method. For our research purpose, we substituted the $l$ in Equations (6) to (7) for $t$. The variable $t$ indicates the time. Then we redefine the Equation (6) as

$$F_d^2(t) = \frac{1}{t}\sum_{n=kt+1}^{(k+1)t}(y(n) - y_l(n))^2, k = 0,1,2,\mathbf{K},(\frac{N}{t}-1) \quad (8)$$

and redefine the Equation (7) as

$$F_d(t) \sim t^a. \quad (9)$$

We will use the new definitions in our work.

### 3.3 Data Analysis

We analyzed the Taiwan stock market for daily data of 3 categories of indices: 1 main index, 19 group indices, and 667 individual stock indices (listed companies). We will use the time series of the main index starting on January 1991 and extending to May 2004 and use the time series of group indices starting on January 1995 and extending to May 2004. The ranges of third category of data are not described here because they listed on the stock market started on different dates, hence we will give a note of each index with date range on the experimental result tables.

The three categories of data are listed as follows:

- Main index: The one index is TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index).
- Group index: The 19 indices are Cement, Food, Plastics, Textiles, Electric Machinery, Appliance Cable, Chemical, Glass Ceramics, Paper Pulp, Steel Iron, Rubber, Automobile, Electron, Construction, Transportation, Tourism, Banking Insurance, Department Stores, and Other.
- Individual stock index: They are listed companies of Taiwan stock market. We don't list them here. They will be listed on experimental result table with their codes (stock codes).

All the collection of data is extracted from the freeware ezChart, which is available on the web [14].

### 3.4 Use of DFA

DFA method was well built as C program by Peng, and this program is free and could be downloaded from reference [15]. After downloading the program, we modified it for our purpose. We rewrote the function input() and added a lsm() function to fit the results from function dfa() and estimate the scaling exponent. Then we used Microsoft Visual C++ to compile this program and generated a DOS program.

The original program has some options and we will give brief description of them as following:

- **-d** *k* : Detrend the data using a polynomial of degree *k* (1: linear, 2: quadratic, etc.). The default value option is linear detrending.
- **-h** : Print a usage summary and exit.
- **-l** *minbox* : Set the smallest box width. The default, and the minimum allowed value for minbox, is 2k + 2 (where k is determined by the -d option, see above).
- **-s** : Perform a sliding window DFA (measure the fluctuations using all possible boxes at each box size). By default, fluctuations are measured using non-overlapping boxes only. Using the -s option will make the calculation much slower.
- **-u** *maxbox*: Set the largest box width. The default, and the maximum allowed value for maxbox, is one-fourth the length of the input series.

Options **-u** and **-l** are both follow a number, if the given numbers are logically conflict (e.g. minbox > maxbox), the program will automatically swap the value of them. And if the given numbers are out of range of the time series, the program will automatically set the default value for them. Because all of our time series data don't have the same number of data points, we use the default values for the two options. As to option **-d** and **-s**, we also use default values for them for the original idea of DFA method.

Our modified program will automatically read the external data described from the database of ezChart and generate the scaling exponent and the time holding the exponent of each time series.

## 4. Experimental Results

In this section, we will give the comparison of the experimental results. Firstly, we will compare the main index with group indices; secondly, we will compare the group index with individual stock indices; thirdly, we will compare the main index with individual stock indices.

### 4.1 Comparison – Main Index vs. Group Index

Table 1 gives the result of the scaling exponent for main index and group index. It can be observed that the main index TAIEX shows a persistent power-law, and 12 group indices also have the persistent correlation, 4 group indices are antipersistent, and others are uncorrelated series. By the results, we found that not all of the group indices follow the same correlation with the main index, but the number of persistent correlation of group indices is grater than half of all group indices. So the results say that more than half of all group indices follow the same correlation with the main index.

**Table 1: Experiment of main and group indices**

| Main Index Name | Scaling Exponent α | Range (trading days) |
|---|---|---|
| TAIEX | 0.53 | 10 → 178 |
| **Group Index Name** | **Scaling Exponent α** | **Range (trading days)** |
| Cement | 0.52 | 10→97 |
| Food | 0.57 | 10→93 |
| Plastics | 0.50 | 10→199 |
| Textiles | 0.55 | 10→197 |
| Electric Machinery | 0.50 | 10→190 |
| Appliance Cable | 0.56 | 10→157 |
| Chemical | 0.51 | 10→120 |
| Glass Ceramics | 0.43 | 10→62 |
| Paper Pulp | 0.53 | 10→142 |
| Steel Iron | 0.51 | 10→235 |
| Rubber | 0.44 | 10→162 |
| Automobile | 0.50 | 10→195 |
| Electron | 0.59 | 10→122 |
| Construction | 0.54 | 10→96 |
| Transportation | 0.45 | 10→200 |
| Tourism | 0.56 | 10→150 |
| Banking Insurance | 0.49 | 10→175 |
| Department Stores | 0.51 | 10→135 |
| Other | 0.53 | 10→136 |

### 4.2 Comparison – Group Index vs. Individual Stock Index

Table 2 gives the result of the scaling exponent for individual indices of Appliance Cable. There are 5 individual indices follow the persistent correlation, 10 individual stock indices follow the antipersistent correlation. Comparing the results with group indices of Appliance Cable (see Table 1), we can observe that all individual stock indices of Appliance Cable don't follow the same correlation with group indices Appliance Cable and there are greater than half of individual stock indices of Appliance following the different correlations from group index of Appliance. This phenomenon can also be found in group indices of Cement, Chemical, Steel Iron, Construction, Tourism, Department Stores and Other with their individual indices (see Table 3). So the results say that there is no obvious relation found between group indices and individual stock indices.

**Table 2: Experiment of individual stock indices of Appliance Cable**

| Stock Code | Scaling Exponent $\alpha$ | Range (trading days) | Starting Date |
|---|---|---|---|
| 1601 | 0.57 | 10→158 | 1991/01/03 |
| 1603 | 0.52 | 10→165 | 1991/01/03 |
| 1604 | 0.46 | 10→230 | 1991/01/03 |
| 1605 | 0.51 | 10→70 | 1991/01/03 |
| 1606 | 0.44 | 10→64 | 1991/01/03 |
| 1608 | 0.49 | 10→44 | 1991/01/03 |
| 1609 | 0.44 | 10→138 | 1991/01/03 |
| 1611 | 0.41 | 10→62 | 1991/01/03 |
| 1612 | 0.48 | 10→236 | 1993/05/08 |
| 1613 | 0.48 | 10→136 | 1995/10/20 |
| 1614 | 0.45 | 10→159 | 1997/10/01 |
| 1615 | 0.55 | 10→29 | 1998/03/31 |
| 1616 | 0.54 | 10→65 | 1997/05/19 |
| 1617 | 0.44 | 10→244 | 1998/04/28 |
| 1618 | 0.42 | 10→37 | 1998/06/23 |

**Table 3: Statistic of results of individual stock indices**

| Group Name | Persistent Correlation | Antipersistent Correlation | Random Correlation | Total |
|---|---|---|---|---|
| Cement | 2 | 6 | 0 | 8 |
| Food | 16 | 8 | 0 | 24 |
| Plastics | 16 | 6 | 0 | 22 |
| Textiles | 25 | 25 | 4 | 54 |
| Electric Machinery | 20 | 15 | 0 | 35 |
| Appliance Cable | 5 | 10 | 0 | 15 |
| Chemical | 8 | 21 | 4 | 33 |
| Glass Ceramics | 3 | 4 | 0 | 7 |
| Paper Pulp | 3 | 2 | 2 | 7 |
| Steel Iron | 9 | 12 | 2 | 23 |
| Rubber | 1 | 8 | 0 | 9 |
| Automobile | 1 | 3 | 0 | 4 |
| Electron | 190 | 69 | 18 | 277 |
| Construction | 11 | 20 | 2 | 33 |
| Transportation | 6 | 11 | 0 | 17 |
| Tourism | 1 | 4 | 1 | 6 |
| Banking Insurance | 19 | 23 | 1 | 43 |
| Department Stores | 3 | 8 | 1 | 12 |
| Other | 16 | 20 | 2 | 38 |
| Total | 355 | 275 | 37 | 667 |

### 4.3 Comparison – Main Index vs. Individual Stock Index

Finally, we compare the correlations between main index and individual indices. There are 355 individual stock indices following the persistent correlation, 275 individual stock indices following the antipersistent correlation and 37 individual stock indices following the random correlation (see Table 4.3). There are greater than half of individual stock indices following the same correlation with main index. So the results say that more than half of individual stock indices follow the same correlation with the main index.

## 5. Conclusions and Future Works

In pervious researches, DFA method is used to detect long-range power-law correlations in main indices. However, we wonder that if the results of DFA method in main indices will imply that group indices and individual stock indices will follow the same correlation with main indices? If not, it is necessary to detect long-range power-law correlations in individual stock indices for long-term investment.

In this paper, we use DFA method to detect long-range power-law correlations in the main index, group indices and individual stock indices and the experimental results show that the long-range power-law correlations exist in the main index as similar as the previous researches and the long memory also exist in most listed companies of Taiwan stock market for the $a \neq 0.5$.

Besides, we found that the correlations detected from main index do not imply the same correlation of group indices and individual stock indices although there are greater than half of group indices and individual stock indices following the same correlation with the main index. We also found that there is no obvious relation between group indices and individual stock indices with their detected correlations.

The DFA method can also substitute quadratic fit or higher order polynomial fit for linear fit $y_t(n)$ (see Equation 8). This will be able to give different views for further researcher to detect long-range correlation in financial time series. Moreover, the persistent type series means that if the past signals have a positive increment, the future signals will be expected as positive and the antipersistent type series means that if the decreasing signals are in the past, the increasing signals will be in the future. According to the tow phenomenon, a forecasting system can be developed while how to implement it will be in the future works.

## References

[1] H. E. Hurst, “Long-term Storage Capacity of Reservoirs,” *Trans. Am. Soc. Civ. Eng. 116*, 1951, pp.770-779.

[2] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, “Mosaic Organization of DNA Nucleotides,” *Phys. Rev. E 49*, 1994, pp.1685-1689.

[3] Detrended Fluctuation Analysis (DFA), http://www.physionet.org/tutorials/fmnc/node5.html

[4] Yanhui Liu, Pierre Cizeau, Martin Meyer, C.-K. Peng, and H. Eugene Stanley, “Correlations in Economic Time Series,” *Physica A 245*, 1997, pp.437-440.

[5] Pilar Grau-Carles, “Long-range Power-law Correlations in Stock Returns,” *Physica A 299*, 2001, pp.521-527.

[6] K. Ivanova, “Toward a phase diagram for stocks”, *Physica A 270*, 1999, pp.567-577.

[7] Robert H. Shumway and David S. Stoffer, *Time Series Analysis and Its Applications*, Springer-Verlag, 2000, Page 1.

[8] T. W. Anderson, *The Statistical Analysis of Time Series*, John Wiley and Sons, Inc., 1971, Page 1.

[9] Sudhakar M. Pandit and Shien-Ming Wu, *Time Series and System Analysis with Applications*, John Wiley and Sons, Inc., 1983, Page 3.

[10] http://reylab.bidmc.harvard.edu/DynaDx/

[11] Rafal Weron, “Estimating long-range dependence: finite sample properties and confidence intervals,” *Physica A 312*, 2002, pp.285-299.

[12] J. Geweke and S. Porter-Hudak, “The Estimation and Application of Long Memory Time Series Models,” *Journal of Time Series Analysis 4*, 1983, pp.221-238.

[13] Kun Hu, Plamen Ch. Ivanov, Zhi Chen, Pedro Carpena and H. Eugene Stanley, “Effect of trends on detrended fluctuation analysis,” *Phys. Rev. E 64*, 2001, 011114, pp.1-19.

[14] http://www.ezchart.com.tw

[15] http://physionet.incor.usp.br/physiotools/dfa/dfa-1.htm