

Chaotic Analysis and Dimension Reduction of Shanghai Composite Index Time Series

Peiqi Zhen^{1, a}, Yue Zeng¹, Jianding Zhou¹ and Lijun Lv

¹School of Beijing Forestry University, Beijing 100083, China.

^a920088054@qq.com

Keywords: Dimension reduction, principal component analysis, chaos theory, Shanghai Composite Index

Abstract. Correctly understanding the fluctuation of the securities market has a great effect on Chinese economic development. In order to verify the chaotic characteristics of Shanghai Composite Index time series, we use G-P algorithm and Cao method to calculate the characteristic quantities such as embedding dimension, correlation dimension and Lyapunov exponent. If the embedding dimension of the time series is large, it will be more complicated to analyze its behavior characteristics. Therefore, we further apply principal component analysis, by using less comprehensive indicators exist in the variables to analyze various information in order to achieve the purpose of dimension reduction.

1. Introduction

Chaos theory is a method that combines quality with quantity, discussing a series of data that cannot be handled solely but should be analyzed and predicted systematically. The object that it studies can be an irregular, aperiodic macroscopic behavior produced by nonlinear dynamic system itself. And the theory can reveal the inherent law of a phenomenon that seems stochastic and disordered. Although Chinese securities market only has nearly 30 years of history, the uneven market often causes social repercussions. Therefore, correctly understanding the chaos law of stock price is very important for our government to strengthen the management of the securities market and to ensure the sustainable development of Chinese economy.

Based on Shanghai Composite Index time series from 2006 to 2010^[1], this paper aims at studying the chaos law and reducing the dimension of Chinese securities market. We respectively use G-P algorithm, Cao method to discuss the chaos mechanism of the evolution and formation of Shanghai Composite Index time series. Then we use principal component analysis to reduce the dimension of the time series. Our research lays a foundation for constructing the prediction model and revealing the inherent rule from a new angle.

2. G-P Algorithm

2.1 Reconstruction of Phase Space

Phase space reconstruction is to reconstruct a chaotic attractor equivalent to the original system through a small amount of information in the known system. Based on this point, Takens Theory^[2,3] was proposed: for a known time series

$$\{S(t), t = 1, 2, \dots, n\} \quad (1)$$

for any delay time τ and embedding dimension m , there must be a certain smooth map

$$F: R^m \rightarrow R^m \quad (2)$$

on the attractor, to meet the condition of

$$Y(t+1) = F[Y(t)], Y(t) = [S(t), S(t+\tau), \dots, S(t+(m-1)\tau)] \quad (3)$$

2.2 Calculate the Correlation Dimension Based on G-P Algorithm

The correlation function^[4] is defined as:

$$C(r) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i \neq j} \theta(r - |x_i - x_j|), \quad \theta(x) = \begin{cases} 1, & (x > 0) \\ 0, & (x \leq 0) \end{cases} \quad (4)$$

In the function, N represents the number of phase points in the phase space. $\theta(x)$ is Heaviside function. For a suitable range of r, the attractor dimension d and the cumulative distribution function C(r) should satisfy the logarithmic linear relationship,

$$d(m) = \ln C(r) / \ln r \quad (5)$$

And the correlation dimension is defined as

$$D_2 = \lim_{r \rightarrow \infty} \ln C(r) / \ln(r_0) \quad (6)$$

Using the above ideas, Fig.1 shows the relationship between correlation dimension and embedding dimension. The correlation dimension tends to be stable and converges to a saturation value $D_2=2.5$. When the embedding dimension of Shanghai Composite Index time series reaches 10, the system has a stable correlation dimension of 2.5. Therefore, Shanghai Composite Index time series has chaotic characteristics.

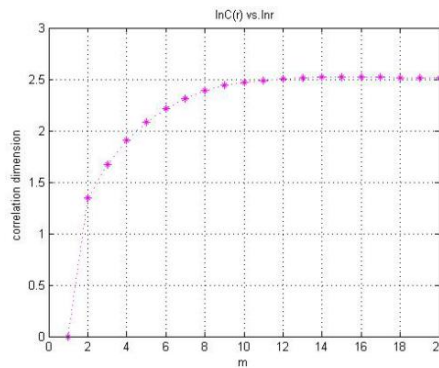


Fig.1 The relationship between correlation dimension and embedding dimension

2.3 Lyapunov Exponent

The Lyapunov exponent^[5] quantitatively describes a phenomenon that the orbits of two very close initial values, separated exponentially over time. It reflects the chaotic level of dynamical system as a whole. Consider two systems,

$$x_{n+1} = f(x_n), y_{n+1} = f(y_n) \quad (7)$$

If there is a small difference in the initial, the two systems will be separated, the degree of separation is measured by Lyapunov exponent, which is the logarithm of the geometric mean

$$\lambda = \frac{1}{n} \ln \left(\prod_{n=0}^{n-1} \left| \frac{df(x_n, \mu)}{dx} \right|_{x_n} \right) \quad (8)$$

Let n be infinite, and get the formula of Lyapunov exponent:

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{n=0}^{n-1} \ln \left| \frac{df(x_n, \mu)}{dx} \right| \quad (9)$$

If $\lambda < 0$, it means that the adjacent points will eventually converge into one point, which corresponds to the stable fixed point and periodic motion. If $\lambda > 0$, it means that the adjacent points will eventually be separated, corresponding to the local instability of the orbit.

Using the above ideas, we take the embedding dimension $m=10$ and calculate the Lyapunov exponent of the Shanghai Composite Index time series. The result is $\lambda_1=0.011$, and the result shows that Shanghai Composite Index time series has chaotic characteristics.

3. Cao Method

3.1 Calculating Embedding Dimension Based on Cao Method.

Cao method is an improvement on FNN method. Cao method defines^[6]

$$a_2(i, d) = \frac{\|X_{d+1}(i) - X_{d+1}^{NN}(i)\|}{\|X_d(i) - X_d^{NN}(i)\|}, E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a_2(i, m), E1(m) = \frac{E(m+1)}{E(m)} \quad (10)$$

In fact, for a finite sequence, it is difficult to determine whether $E1(m)$ is changing slowly or is already stable, so supplements the criterion

$$E^*(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} |X(i + m\tau) - X^{NN}(i + m\tau)|, E2(m) = \frac{E^*(m+1)}{E^*(m)} \quad (11)$$

For deterministic sequences, $E1(m)$ will converge to a certain value after m is greater than a certain value m_0 and there are always some m values that make the value of $E2(m)$ fluctuates around 1.

According to the above ideas, we use Cao method to calculate the embedding dimension of Shanghai Composite Index time series. The calculated result is $m=9$. Fig.2 shows that the obtained $E1$ is stable when the embedding dimension $m=9$. There are some fluctuations in $E2$, but the volatility is not big, indicating that Shanghai Composite Index time series is more random.

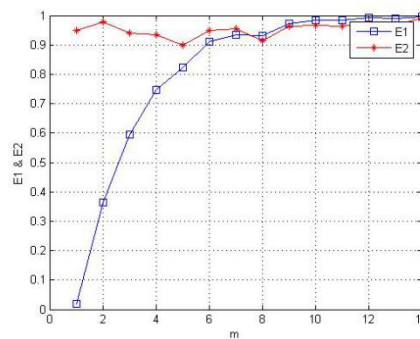


Fig.2 E1&E2 changing with embedding dimension m

3.2 Calculate Lyapunov Exponent

We take the embedding dimension $m=9$ calculated by Cao method and calculate the Lyapunov exponent. The result is $\lambda_2=0.0141$, and it is not quite different from $\lambda_1=0.011$, which indicates that the improvement of Cao method is effective.

4. Principal Component Analysis

4.1 The Model of Principal Component Analysis.

The stock market is risky and uncertain to some extent, so the analysis and prediction of the stock market is very important. If the Shanghai Composite Index time series has a high dimension, it will be very difficult to analyze and forecast. Reducing the dimension properly can improve the prediction speed and lower the risks. Principal component analysis aims to combine and synthesize relevant information in real life. It is an effective way to simplify the data structure through dimension reduction. Therefore, we can use less comprehensive indicators to analyze various types of information under the principle that the loss of data is the least. Obviously, it is easier to analyze the data series in a low-dimensional space.

4.2 Dimension Reduction.

We use principal component analysis to reduce the dimension of Shanghai Composite Index time series. For the variables with correlation, we merge them by using averaging method, and the calculated embedding dimension is $m=6$. Calculate Lyapunov exponent of the time series with $m = 6$, and the result is $\lambda_3=0.0182$, which is not much different from λ_1 and λ_2 . The result shows that we use principal component analysis to achieve the purpose of dimension reduction, and retain the information of the original time series as much as possible.

5. Conclusions

According to the above analysis, we conclude that Shanghai Composite Index time series has chaotic characteristics by calculating characteristic quantities. Using G-P algorithm and Cao method to calculate embedding dimension, and use the results to calculate Lyapunov exponent. Because the two results are similar, we conclude that the improvement of Cao method is effective. To facilitate the analysis and prediction, we reduce the dimension of the time series by using principal component analysis. In addition, how to forecast the Shanghai Composite Index time series effectively still need to be further studied.

6. Acknowledgments

This project was supported by Beijing Municipal Training Program of Innovation and Entrepreneurship for Undergraduates. (NO: S201610022104)

7. References

- [1] Information on <http://q.stock.sohu.com/zs/000001/lshq.shtml>
- [2] Jinhua Lv, Junan Lu, Shihua Chen, Chaotic time series analysis and its application, Wuhan, 2002.
- [3] Yunzhong Liu, Huiyu Xuan, Chaotic Time Series and Its Application of GDP(1978~2000) Forecasting in China, Journal of Industrial Engineering/Engineering Management. 1004-6062(2004)02-0008-03
- [4] Baohui Men, Xiejing Zhao, Chuan Liang, Chaos analysis on precipitation time serial in Sichuan middle region of Yangtze River upstream reaches, Journal of Yangtze River Scientific Research Institute. 1001-5485(2004)01-0043-04
- [5] Information on <http://baike.so.com/doc/4143005-4342724.html>
- [6] Cao L Y, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D: Nonlinear Phenomena. 1997, 110(1-2):43-50.