Physica A 473 (2017) 632–639

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Correlation dimension of financial market

Chun-Xiao Nie

Department of Finance, School of Business, East China University of Science and Technology, Shanghai 200237, China

HIGHLIGHTS

- The correlation dimension is applied to analyze financial markets.
- The financial crisis led to dramatic changes in the correlation dimension.
- The correlation dimension of the simulation data based on geometric Brownian motion is significantly larger than that of the real market.

ARTICLE INFO

Article history:
Received 13 October 2016
Received in revised form 19 December 2016
Available online 6 January 2017

Keywords:
Dimension
Financial market
Time series
Financial crisis

ABSTRACT

In this paper, correlation dimension is applied to financial data analysis. We calculate the correlation dimensions of some real market data and find that the dimensions are significantly smaller than those of the simulation data based on geometric Brownian motion. Based on the analysis of the Chinese and US stock market data, the main results are as follows. First, by calculating three data sets for the Chinese and US market, we find that large market volatility leads to a significant decrease in the dimensions. Second, based on 5-min stock price data, we find that the Chinese market dimension is significantly larger than the US market; this shows a significant difference between the two markets for high frequency data. Third, we randomly extract stocks from a stock set and calculate the correlation dimensions, and find that the average value of these dimensions is close to the dimension of the original set. In addition, we analyse the intuitional meaning of the relevant dimensions used in this paper, which are directly related to the average degree of the financial threshold network. The dimension measures the speed of the average degree that varies with the threshold value. A smaller dimension means that the rate of change is slower.

1. Introduction

The correlation dimension is introduced in chaos theory, and is used as a type of fractal dimension to measure the set of random points in space [1–3]. Some methods have been proposed for estimating the correlation dimension [4,5]. Correlation dimensions have been widely used in time series analysis such as [6–8].

Because there are a large number of time series in the financial market, the correlation dimension of some financial time series is also estimated [9–11]. The authors have used the correlation dimension to analyze the high frequency data of the stock and found that the real system is high-dimensional or low-dimensional, but high entropy that cannot be used for prediction [9]. In addition, in [10], the researchers studied the data in the currency market, the correlation dimension was applied, and found that there is a nonlinear structure in the Dllar/Pound daily rate. The correlation dimension has also

E-mail address: niechunxiao2009@163.com.

http://dx.doi.org/10.1016/j.physa.2017.01.045

been used to analyze the Chinese stock market data, the authors found that some stock time series to show the chaotic behavior [11].

In addition to time series analysis, correlation dimensions are also used for machine learning, which is an interdisciplinary for exploring the study and construction of algorithms that can learn from and make predictions on data. The correlation dimension is applied to estimate the intrinsic dimension of data sets [12,13]. From an abstract perspective, the correlation dimension is applied to a finite set of metric space [13]. As a special case, Euclidean space is a special and commonly used metric space.

The central idea in this paper is to calculate the correlation dimension of a finite set of stocks, which is a discrete metric space when the appropriate metric is defined on the set. Moreover, in addition to the Euclidean distance, the metric proposed in [14] has been widely used in recent literature, and hence this paper, too, will use this metric.

In our approach, the distance of a set of stocks is calculated first, and then the correlation dimension is calculated. In the definition of the correlation dimension, the Heaviside step function is used [1–3]. It specifies a certain threshold and the distance matrix is transformed into a 0–1 matrix. At the same time, the 0–1 matrix can be regarded as an adjacency matrix of a threshold network.

In fact, recently, the threshold method has also been used in the study of financial networks. In previous studies, the threshold method has been applied to American, Chinese, and Korean stock markets [15–17] and to market indices [17,18]. Some researchers have found that the power law exists in threshold networks [15–17,19–23], especially in financial networks during a financial crisis [16,17]. In summary, this paper can be considered as an extension of the threshold network research.

2. Data and methods

2.1. Data

The five data sets used in the Chinese and US markets are denoted as D_1, D_2, D_3, D_4, and D_5, respectively.

The daily closing price series of S&P500 constituent stocks, from 2005/1/3 to 2014/9/3, amounting to a total of 2436 trading days, was selected. Stocks with missing data points were removed and 448 stocks from the S&P500 constituent stocks were used (Data D_1).

The daily closing price series of CSI300 constituent stocks, from 2005/1/4 to 2009/12/31, was also used. Stocks with missing data points were removed and 246 stocks were used (Data D_2).

Furthermore, the 5-min price series of SSE180 constituent stocks was also used. Data from 2013/8/13/11:05 to 2013/8/19/15:00 was selected. This amounted to a total of 180 stocks and the length of the time series was 222×5 minutes (Data D_3).

For comparing the US and Chinese stock markets, the 5-min price series of S&P500 constitute stocks, from 2015/6/1 to 2015/12/22, and CSI300 constituent stocks, from December 2015/6/1 to 2015/12/20, were selected (Data D_4).

Finally, for further analysis of the financial crisis, the daily closing price data for the S&P500 constituent stocks from January 3, 1985 to December 31, 1988, was used. As many companies had missing price data, only 78 stocks were selected (Data D_5).

2.2. Methods

We assume that the point set $V = \{v_i, i = 1 \dots n\}$ is in a metric space, and C_n is defined as shown in Eq. (1), where θ is the Heaviside function and r is the threshold, and $d(i, j)$ is the distance between v_i and v_j. The correlation dimension (Dim) is then estimated by plotting $\log[C_n(r)]$ against $\log(r)$ and estimating the slope of the linear part of Eq. (2).

In this paper, we will discuss the correlation dimension of a financial market. Therefore, we first need to define the elements of set V and metric. In our study, $\{v_i\}$ is the stock where each stock corresponds to a price series $\{P_i(t)\}$. The distance between the stocks is calculated based on the logarithmic return series (Eq. (3)). The Pearson correlation coefficient between the stocks is shown in Eq. (4). We use the definition originally proposed by [14] to define the distance between stocks (Eq. (5)), which satisfies the three conditions of the metric definition.

Based on the definition of stock distance, we introduce the following calculation method, which consists of the following four steps.

Step 1: We calculate the distance matrix $D = [d_{ij}]$ based on Eqs. (4) and (5).

Step 2: We calculate the upper triangular matrix of D and sort the elements in the set $\{d(i, j), j > i\}$. Two parameters p_1 and p_2 are set and named threshold ratio, which correspond to the two threshold values e_1 and e_2, respectively. The proportion of elements smaller than e_1 is p_1, and the proportion of edge weight smaller than e_2 is p_2. If we calculate the correlation dimension based on the sliding window, we do not need to set parameters repeatedly. This step makes it easy to repeat large numbers of calculations. We will discuss the problem of parameter selection in Section 3.1.

Step 3: We select some elements from the interval $[e_1, e_2]$ as the threshold and denote $\{r_k\}$ to calculate the value of the correlation function $C_n(r_k)$.

Step 4: In the double logarithmic coordinate system, we estimate the Eq. (2) using the least squares method, where the coefficient Dim is the dimension.