Abstract

We investigate the statistical properties of traders’ trading behavior using cumulative distribution function (CDF). We analyze exchange data of 52 stocks for one-year period which contains non-manipulated stocks and manipulated stocks published by China Securities Regulatory Commission (CSRC). By analyzing the total number of transactions and the trading volume of each trader over a year, we find the cumulative distributions have power-law tails and the distributions between non-manipulated stocks and manipulated stocks are different. These findings can help us to detect the manipulated stocks.

Keywords: stock market, trading activity, manipulation detection

1. Introduction

In recent years, the dynamic behavior of stock market has attracted much attention. Based on large amounts of historical data, the dynamic behavior of stock market has been analyzed in understanding of the stock markets. Market manipulation is a universal phenomenon, and it is very difficult to supervise because the methods used to manipulate are underground. Manipulation may distort the stock-prices and is harmful to the stock market. Since manipulation is a very important issue and it affects the investor confidence, it is of importance to detect, deter and prevent it. Manipulation has drawn the attention of many researchers.

The first researchers who studied manipulation are Allen, Gale and Jarrow[1, 2]. They studied the history of the stock-price manipulation and classified the manipulations as three types: action-based manipulation, information-based manipulation and trade-based manipulation. Today, the most common types of manipulation are information-based manipulation and trade-based manipulation and many investigations are about these two types of manipulation. Felixon and Pelli examined the closing-price manipulation[3]. Aggarwal and Wu developed a model to explain
trade-based manipulation and tested the model by using data from US stock markets[4]. Few researchers have attempted to detect manipulation. Pirrong used regression analysis and error correction models to detect manipulation in futures market[5]. Öğüt and Doğanay use data mining techniques (Artificial Neural Network(ANN) and Support Vector Machine(SVM)) to detect stock-price manipulation of Turkey[6].

Traditional researches on the dynamic behaviors of stock market are using probability distribution functions [7, 8, 9, 10, 11, 12, 13] and correlation functions[14, 15] to study the dynamic behaviors of stock markets at the macro level and find the power-law distribution of trading volumes[7, 8, 11, 12, 16, 17]. DFA(Detrended Fluctuation Analysis) has been considered to be better than other methods in measuring correlation and fractality[18]. Based on DFA, Matia and Ashkenazy find that the price fluctuations of stocks showed multifractal properties[19]. Many researchers use these approaches to study the dynamical behavior of the Chinese stock market by investigating the statistical properties[20, 21, 22, 23] and analyzing the multifractality[24, 25]. These studies investigate the dynamic behavior of stock markets at the macro level. However, the participants of stock market are listed companies, organizations and a mount of investors. Our work focus on traders’ behavior and using the exchange data instead of stocks’ indices.

In trade-based manipulation, manipulators engage in fraudulent trading to create an image of an active market and attract the other investors to buy the stock[1]. A group of manipulators have "heavy trading" among themselves as compared to their trading with others. Therefore, we investigate the dynamic behavior of stock markets based on details to find some statistical properties between non-manipulated stocks and manipulated stocks. We analyze two microscopic quantities: the total number of transactions $N_T$ and the total number of trading volume $Q_T$ of each trader for one stock. Our analysis shows that in the non-manipulated stocks the distributions of $N_T$ displays power-law decay, and the distributions of $Q_T$ has power-law tails with the exponents $\beta$ between 0.75 to 1.05. Furthermore, we compare the non-manipulated stocks and manipulated stocks, and we find the distributions of $N_T$ and $Q_T$ are different. Finally, we take the difference of the distributions to detect the manipulated stocks. Our study differs from other studies in this literature at least one of the following reasons. First, we investigate the dynamic behavior of stock markets at the micro level. Second, we focus on detecting stock price trade-based manipulation, and we use real world exchange data instead of synthetic database.

This paper is organized as follows: Section 2 presents the exchange data analysis and distributions of $P(N_T)$ and $P(Q_T)$. Then, we compare the normal exchange data and the anomalous data. In section 3, a simple method based on previous analysis of manipulation detection is proposed, while conclusions are presented in Section 4.

2. Analysis of trading activity

2.1. Dataset

Our empirical results are based on the analysis of a dataset covering every transaction record of 52 securities in 2004. The total number of transactions for 52 securities is of the order of 12,585,747 in one-year period studied. There are 3,689,435 traders in this dataset and average number of traders of each stock is 70,950. These 52 securities contain 45 non-manipulated stocks and 7 manipulated stocks. 4 of the 7 stocks have the manipulated period through a whole year and the rest 3 stocks’ manipulated period is from Jan 2004 to Sep 2004. Because the capital stocks possibility affect the traders behavior, we analyze the exchange data by dividing the total stocks into three subsets based on the capital stocks: 170–250, 80–110 and 40–50 million shares.
We analyze the number of transactions of each trader for one stock in a year to find the statistical properties of traders exchange behavior. First, we consider the number of transactions \( N_T \) of each trader for one stock in one year period. Figure 1 shows, for three types of normal stocks, the cumulative distributions \( P(N_T > x) \) which are consistent with a power-law decay,

\[
P(N_T > x) \sim (N_T)^{-\alpha}
\]

(1)

From Figure 1, we observe the power-law distribution at the tails. Using the least-square fit, we obtain the power-law exponents \( \alpha \) for three subsets.

2.3. The distribution of share-volume

In order to understand the total share-volume traded of one trader in a year for one stock, we analyze the share-volume of each transaction for one trader. We note that

\[
Q_T = \sum q_T^i,
\]

(2)

is the sum of the number of shares traded of each trader \( T \) for a given stock, where \( q_T^i \) is trader \( T \)'s volume of every transaction. Hence, we next analyze the statistical properties of the total share-volume of one trader for a given stock in a year and find that the distribution \( P(Q_T > x) \) displays a power-law decay,

\[
P(Q_T > x) \sim (Q_T)^{-\beta}.
\]

(3)

From Figure 2, we observe the power-law distribution at the tails and the exponent \( \beta \) almost similar. Using the least-square fit, we obtain the power-law exponents \( \beta \) for 45 non-manipulated stocks shows an approximately Gaussian spread around the average value \( \beta = 0.92 \pm 0.16 \).

2.4. Trading activity of manipulated stocks

Prices of the stocks must be determined by the listed companies’ condition without any interference, and the price of a certain stock can affect the investors’ decision to buy or to sell it. When the price of a stock begins to rise, some investors buy the stock. While the price begins to reduce, some investors sell it. Therefore manipulators employ different methods to influence the price of the targeted stocks. In trade-based manipulation, manipulators engage in fraudulent trading to create an image of an active market and attract the other investors to buy the stock[1].

In order to understand the manipulated stocks, we analyze 7 manipulated stocks published by China Securities Regulatory Commission(CSRC). These 7 manipulated stocks contain one capital stocks in 170–250, one in 40–50 and five in 80–110. 4 of the 7 stocks have the manipulated period through a whole year and the rest 3 stocks’ manipulated periods are from Jan 2004 to Sep 2004. For the manipulated stocks, we compute the distributions of \( P(N_T) \) and \( P(Q_T) \) for the manipulation period. We also compute the same statistics of the non-manipulated stocks for the same periods.

Figure 1 shows the distribution of \( P(N_T) \) of the three subsets and Figure 2 shows the distribution of \( P(Q_T) \). Comparing to non-manipulated stocks and manipulated stocks, the difference is obvious. The decline of the curves of manipulated stocks are slower than the non-manipulated stocks'. In fact, manipulators trade among themselves in order to artificially increase the price
Figure 1: (a) Cumulative distribution function $P(N_T)$ on a log-log scale for 13 stocks and 1 manipulated stock show a power-law behavior characterized by an exponent $\alpha=1.53$, and the capital stocks are in $170 \sim 250$. (b) 14 non-manipulated stocks and 5 manipulated stocks’ capital stocks are between 80 to 110 and $\alpha=1.55$. (c) 18 non-manipulated stocks and 1 manipulated stock’s capital stocks are between 40 to 50 and $\alpha=1.68$. 
Figure 2: (a) Cumulative distribution function $P(Q_T)$ for shows a power-law tail characterized by an exponent $\beta = 0.91$, and the capital stocks are in 170 - 250. (b) The similar case of capital stocks between 80 to 110 and $\beta = 0.904$. (c) The similar case of capital stocks are between 40 to 50 and $\beta = 0.94$. 
and volume of a stock for the purpose of attracting the other investors to buy the stock and they have heavy trading among themselves. Manipulators earn a profit and investors incur losses. The anomaly in these distributions means the manipulators are disturbing the market regular.

After analyzing these manipulated stocks, we find they have much fewer holders than the other non-manipulated stocks in the same capital stocks. It means a minority of holders has a great part of shares.

By the analysis of traders transactions, we find: (i) Both the distributions of $P(N_T)$ and $P(Q_T)$ are follow power-law at the tails. (ii) Manipulated stocks show obvious differences in $P(N_T)$ and $P(Q_T)$. (iii) The curves of manipulated stocks deviate from the non-manipulated stocks’.

3. Manipulation detection in stock exchange

From the analysis of above section, we can discriminate between non-manipulated stocks and manipulated stocks by computing the diversification of all distribution of $P(N_T)$.

First we compute the distributions of $P(N_T)$ for all stocks and sort all distributions by curves’ position from top to bottom and remove the scatters in the tails. Second choose n position $x_i$ randomly and compute the diversification of each curve. We define the diversification of two different curves i and j:

$$d_{ij} = \sqrt{\sum_{k=1}^{k=n} [p_i(x_k) - p_j(x_k)]^2}, \ (i < j)$$

where $p_i(x_k)$ and $p_j(x_k)$ are value at $x_k$ in the curve i and curve j. Then, we sort the maximum diversification of each stock i in descending order. We find the diversification between the manipulated stocks and non-manipulated stocks is evident, while the diversification of non-manipulated stocks follows a linear function. By using least-square fit the linear function $y=a+bx$, the top largest points above the line are the manipulated stocks with great probability.

Figure 4 is an example for capital stocks in 80 ~ 110. The five points of largest value above the line are the anomalous stocks in our sample. By analyzing the top two manipulated stocks,
we find they have much fewer holders than the other non-manipulated stocks in the same capital stocks.

4. Conclusion

The aim of this research is to detect trade-based manipulation by computing the diversifications of all stocks’ distributions. For this purpose, we analyze two distributions: the total number of transactions of each trader and total share-volume traded of each trader, and we compare these two distributions between non-manipulated stocks and manipulated stocks. We find that using exchange data to detect anomaly is easier than the stock’s indices and propose a simple method to detect anomaly. Finding the collaborated manipulators is the future work.


