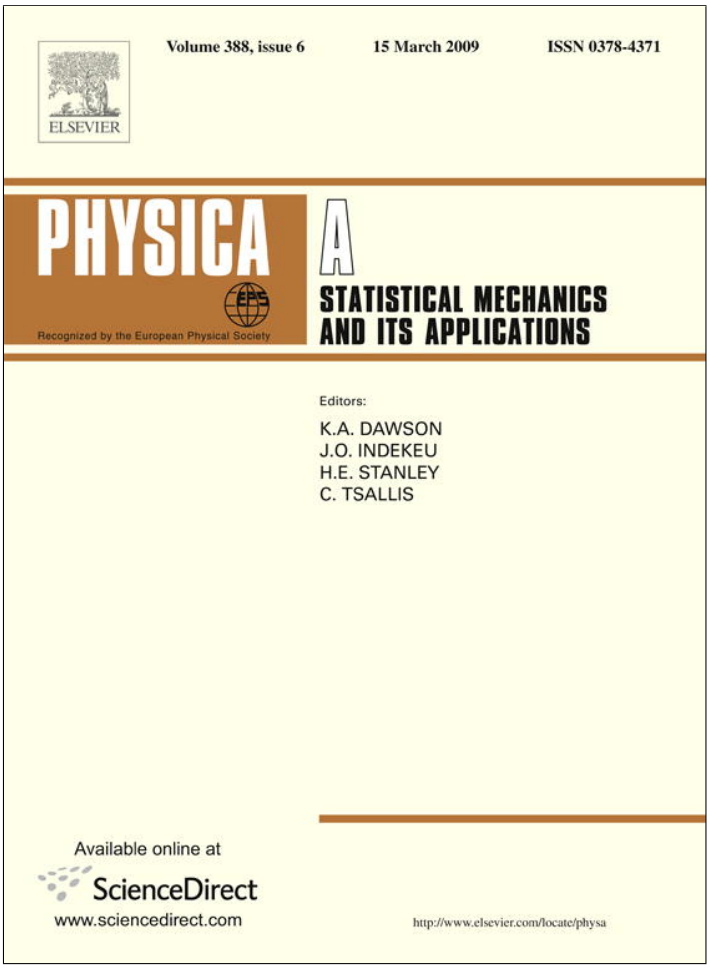

Physica A 388 (2009) 891–899

Contents lists available at ScienceDirect

Physica A

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

Peculiar statistical properties of Chinese stock indices in bull and bear market phases

W.C. Zhou, H.C. Xu, Z.Y. Cai, J.R. Wei, X.Y. Zhu, W. Wang, L. Zhao, J.P. Huang *

Department of Physics and Surface Physics Laboratory (National Key Laboratory), Fudan University, Shanghai, 200433, China

ARTICLE INFO

Article history:
Received 24 June 2008
Received in revised form 30 October 2008
Available online 28 November 2008

PACS:
89.65.Gh
89.75.Da
02.50.-r

Keywords:
Chinese stock market
Bull and bear markets
Price's log-return

ABSTRACT

Chinese stock markets have experienced an extraordinary bull market since Jan 2006, which attracted global eyes. We investigate the statistical properties of the indices' log-return $r(t)$ for the bull market (Jan 2006–Oct 2007) and the previous bear market (Jan 2001–Dec 2005). Here we report three peculiar features of $r(t)$: (i) the cumulative distribution function curve of $r(t)$ in the bull market is similar to that in the bear market; (ii) the autocorrelation function of $r(t)$ in the bull market has a stronger negative correlation and a shorter correlation time than that in the bear market; (iii) the bull market shows stronger long-term correlation than the bear market. This work has relevance to understanding novel statistical properties in economic systems.

1. Introduction

The distribution function of stock price's log-return is an issue that has interested researchers for a long time. Bachelier stated that market price would follow a random walk and the probability distribution $P(r)$ of price return $r(t)$ would be Gaussian. 60 years later, Mandelbrot argued that this was wrong in his research for cotton markets. He substituted the Gaussian distribution with a Levy-stable one, in which the tail decay is weaker than $1/r^3$ [1]. And in 1995, Mantegna and Stanley refuted the Levy-stable distribution with a power-law decay, in which the tails approximately obey $1/r^4$ [2]. This result was later verified by many other researchers on different markets [3–10].

Inspired by Mantegna and Stanley's work, physicists began studying the statistical properties of financial market with methods widely used in statistical physics. Even methods in complex systems have been used, resulting in the discovery of many "universal" rules [2–6]. Among all those rules, the behavior of autocorrelation functions of $r(t)$ and $|r(t)|$ is one of the most important findings. $r(t)$ is reported to have short-term autocorrelation, while $|r(t)|$ is long-term correlated.

However, all these "universal" rules were challenged when people extended their analysis into newly-born markets. The stock price's log-return in India shows exponential distribution function instead of the power-law decay [11]. Additionally, Huang found that the HangSeng index of the Hong Kong stock market shows an exponential decay when skipping the data in the first 20 min of every trading day [12].

As another newly-born market, the Chinese stock markets have attracted researchers, who wish to testify the ubiquity of the power-law decay and other "universal" rules. Recent studies reported that the Chinese stock markets show many similar behaviors as mature markets [13–15].

The previous studies focused on difference among regions. In this work, however, we want to test the power-law decay and the correlation behaviors of $r(t)$ under different market phases, the bull market and the bear market. As is known, the

* Corresponding author. Tel.: +86 21 5566 5227.
E-mail address: jphuang@fudan.edu.cn (J.P. Huang).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.11.028

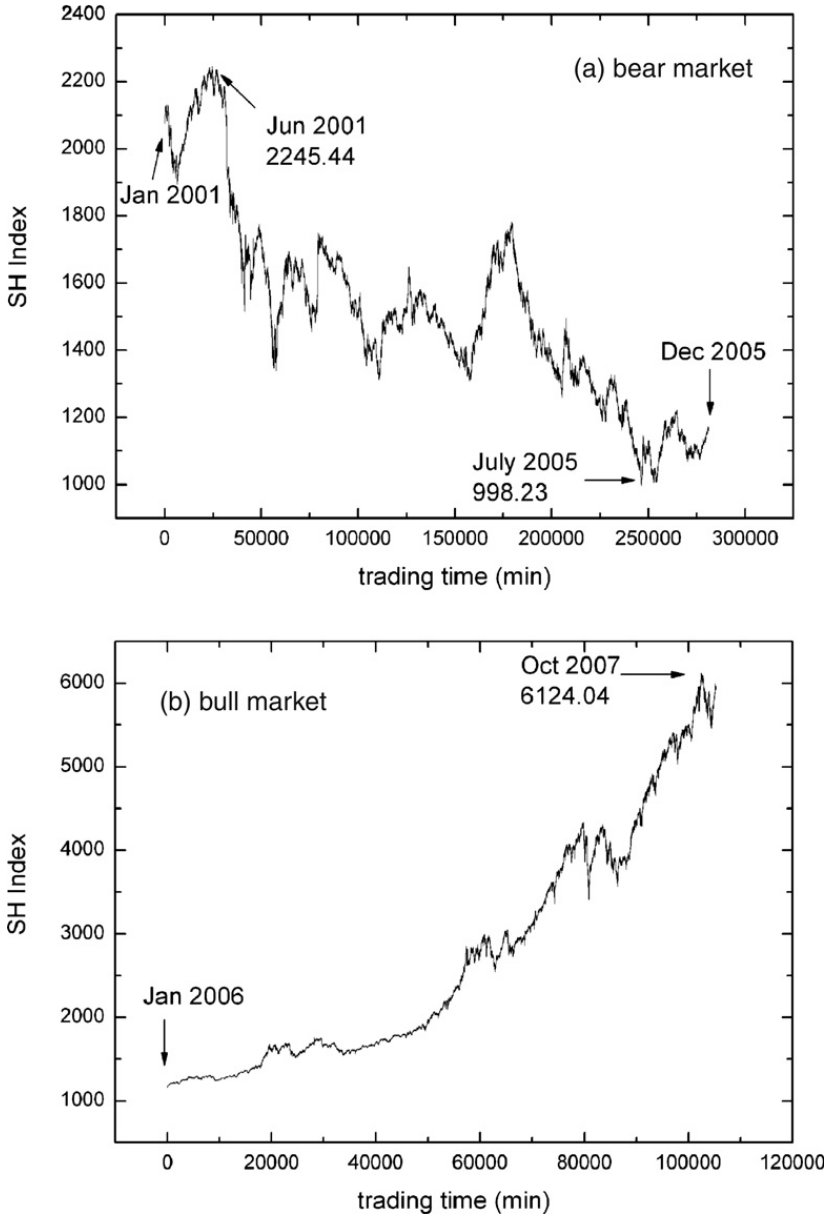

Fig. 1. The evolution of SH Index from Jan 2001 to Oct 2007, which can be divided into two different market phases, (a) the bear market and (b) the bull market. The bear market lasted from about Jan 2001 to Dec 2005, and the bull one lasts from Jan 2006 to October 2007.

Chinese stock markets experienced a nearly 5-year-long bear market and turned its direction in the end of 2005. Take the SH Index for an example, in October 2001, it reached a peak at 2245.44 and dropped to a trough at 998.23 in July 2005 (Fig. 1(a)). After that, the SH Index rose dramatically and reached a new peak at 6124.04 in 19 October 2007 (Fig. 1 (b)). During this period, the total market capitalization of Chinese stock markets has become over five times as much as that of two years ago.

Though the bear market contains rebound period and the bull market includes decline period, this division about the bear and the bull markets is appropriate due to the following reasons. First, this division is in accordance with conventions in finance industry. Most professionals viewed the history this way. Second, what determine the properties of the stock market are the participants. From 2001 to 2005, Chinese stock investors were mainly distressed by the endless decline of the stock market. Even though several rebounds occurred, the market confidence was never completely recovered. People in the bear market became conservative and the trading volume was very small. However, after the year 2006, the Chinese stock market turned into the bull market due to a series of favorable news and policies. The market players became enthusiastic, which could be reflected by the huge trading volume (Fig. 6). We believe that the different statuses of the market participants determine the features of the bull and the bear markets. So this division about the two market phases can be taken in this work.

To sum up, this article's aim is to study the stock market from a new perspective, to find out the similarity and difference between the bear and the bull markets. In Section 2, we will introduce the database we use. In Section 3, we will show the similarity between the distributions of $r(t)$ in different market phases. The differences that lie in the correlation functions in the two market phases will be discussed in Sections 4 and 5.

Table 1

Index	Bear market/the number of records	Bull market/the number of records
SH Index	From Jan 2001 to Dec 2005	From Jan 2006 to Oct 2007
SZ Index	2.8×10^5 records	1.0×10^5 records
CSI-300	None	

2. Data analyzed

There are two stock exchanges in China, the Shanghai Stock Exchange and the Shenzhen Stock Exchange. We analyze three important indices in both stock exchanges. First, we analyze the Shanghai Composite Index (SH Index), which comprises all listed stocks at the Shanghai Stock Exchange; and the Shenzhen Component Index (SZ Index), which consists of 40 major listed stocks at the Shenzhen Stock Exchange. Our database for both SH Index and SZ Index covers the period from Jan 2001 to Oct 2007, with a recording frequency of 1/min. Second, we study the China Securities Index 300 (CSI-300), which is composed of 300 major listed stocks at both the Shanghai and Shenzhen Stock Exchanges. The database for CSI-300 covers the period from Apr 2005, when this index was launched, to Oct 2007. The recording frequency for CSI-300 is also 1/min. All the above indices are market-capitalization weighted indices. The constituents for the SZ Index and CSI-300 are to be adjusted periodically.

In order to compare the statistical properties of Chinese stock indices in different market phases, we divide our data into two parts: the period from Jan 2001 to Dec 2005 is the bear market; the period from Jan 2006 to Oct 2007, the bull market. Because CSI-300 was launched in Apr 2005, we only analyze its bull market part. Table 1 is the snapshot of our database. All the data were bought from www.wstock.net. We have excluded the wrong points which are caused by recording errors such as misprint of the digit.

3. The distribution in different market phases

We define the price (log-)return of time interval Δt as

$$r_{\Delta t}(t) \equiv \ln p(t + \Delta t) - \ln p(t), \tag{3.1}$$

where $p(t)$ is the price of a stock and Δt is the time interval for calculation. In order to compare the results with standard normal distribution $N(0, 1)$, we define the normalized return as

$$g_{\Delta t}(t) \equiv \frac{r_{\Delta t}(t) - \langle r_{\Delta t} \rangle}{\sigma(r_{\Delta t})}, \tag{3.2}$$

where $\langle x \rangle$ denotes the average of x and $\sigma(x)$ is the standard deviation. Here we remove the nights and holidays from the data set so that everyday closing price is followed by the next morning opening price. Since Δt is fixed at 1 min in this article, we will omit the subscript Δt in the following; that is, we will rewrite $r_{\Delta t}(t)$ as $r(t)$, $g_{\Delta t}(t)$ as $g(t)$.

Now we calculate the cumulative distribution function (CDF) of $g(t)$ for the SH Index and SZ Index under the two market phases respectively. The differences that lie in the tails of the CDF curves are too subtle to be distinguished (Fig. 2(a) and (b)). We consider that these differences are caused by the intrinsic errors of the stock market. So we put the five data sets together, the bull and bear market data of SH Index and SZ Index and the bull market data of CSI-300, in order to calculate the CDF with error bars (Fig. 2(c) and (d)). The tails of CDF curve show conspicuous power-law behavior:

$$P(|g| > x) \sim x^{-\alpha}, \quad \alpha = \begin{cases} 2.29 \pm 0.05, & \text{positive tail} \\ 2.18 \pm 0.04, & \text{negative tail.} \end{cases} \tag{3.3}$$

To sum up, the different market phases have little impact on the distribution curves of price's return. The tails of distribution curves decay as power-law.

4. Autocorrelation function

The autocorrelation function is another important feature of the index log-return. If $B(t)$ is a time series, then the autocorrelation function is defined as

$$\rho(\tau) = \frac{E\left[(B_t - E(B_t))(B_{t+\tau} - E(B_{t+\tau}))\right]}{\sqrt{D(B_t)D(B_{t+\tau})}}, \tag{4.1}$$

where B_t is $B(t)$ with the last τ elements removed and $B_{t+\tau}$ is $B(t)$ with the first τ elements removed. In the above equation, $E(x)$ is the expectation of x, and $D(x)$ is the standard deviation.

In this article, we first calculate the autocorrelation function of index log-return $r(t)$ (Fig. 3). The result is interesting for two aspects. On the one hand, we confirm the existence of the negative correlation, which shows that the Chinese

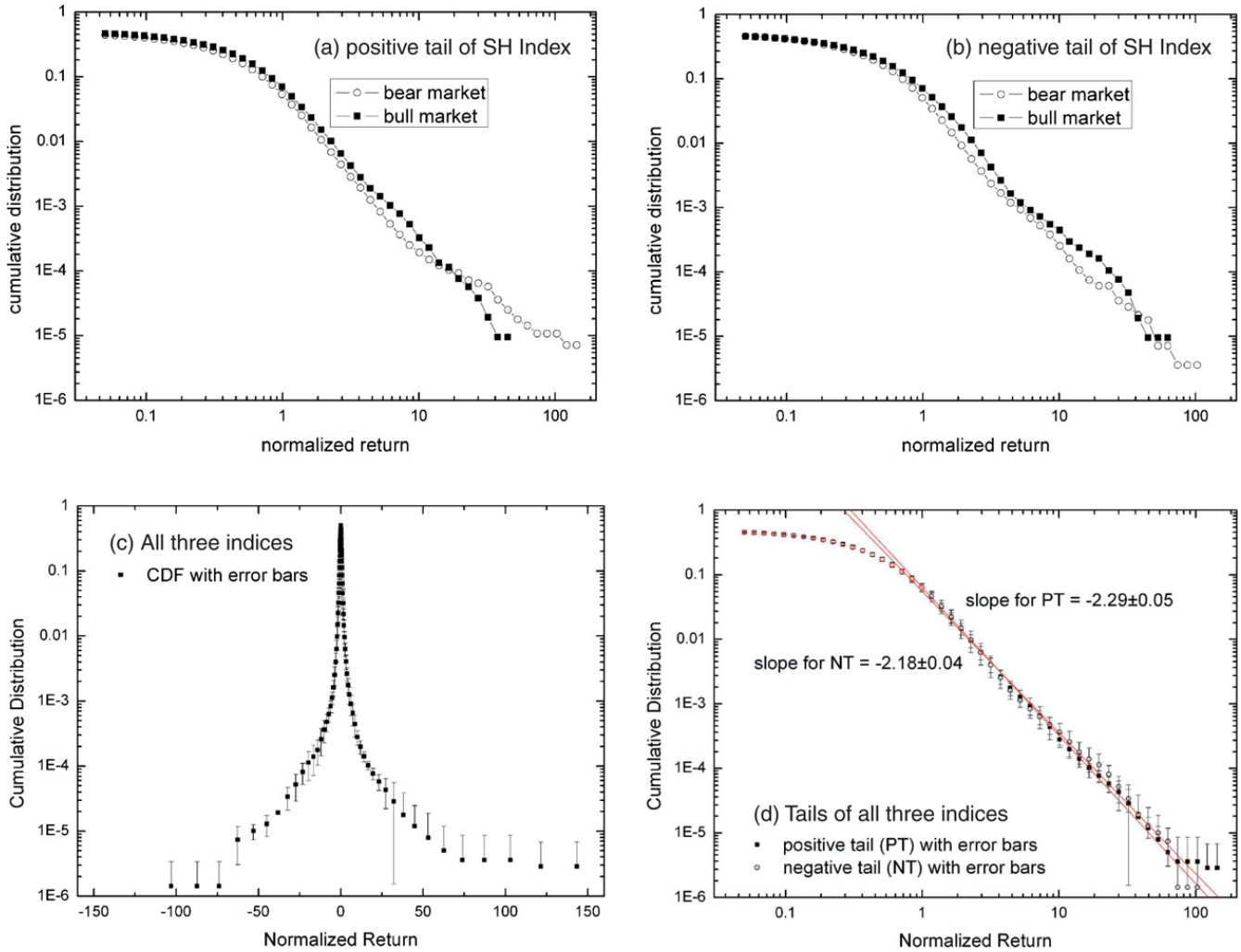

Fig. 2. The (a) positive tail and (b) negative tail of cumulative distribution function (CDF) of normalized return for SH Index under the two market phases respectively. The bear market is displayed in circle, while the bull one is in square both in (a) and (b). We omit the CDF tails for SZ Index since they are similar to those for SH Index. The plots show that the CDF curve keeps its form in different market phases. So we put the five data sets of all the three indices together and calculate the CDF curve and CDF tails with error bars in (c) and (d) respectively.

market is similar to western developed markets (as described in Ref. [5]). On the other hand, we find that the autocorrelation curve of the bull market has a stronger negative correlation and a shorter correlation time (a time period through which the autocorrelation function falls into the noise level) than that of the bear market. We will give our explanation for this phenomenon in Section 6.

Second, we compute the autocorrelation for the absolute log-return $|r(t)|$ (Fig. 4), in which the intraday pattern is removed (the intraday pattern is discussed in the Appendix). As is reported [3,4,7,15], the autocorrelation function of $|r(t)|$ obeys a power-law decay. In Fig. 4, we see this power-law decay clearly. We also find in the figures that the bull market has a larger value of autocorrelation function for $|r(t)|$ and need longer time to decay to the noise level. Thus, the bull market has a stronger long-term autocorrelation than bear market. This will be demonstrated more conspicuously with the DFA method in Section 5.

5. DFA analysis

For a more precise study in the long-term autocorrelation of $|r(t)|$, we employ the detrended fluctuation analysis (DFA), which was proposed more than a decade ago [16,17] and has been successfully applied to analyze long-range autocorrelation in various systems.

Here we apply the DFA method to the volatility time series $|r(t)|$ with the intraday pattern removed (The intraday pattern is discussed in Appendix). To calculate the DFA function, we should first construct

$$C(t') = \sum_{t''=1}^{t'} \left[\left| r(t'') \right| - \langle |r(t)| \rangle \right], \tag{5.1}$$

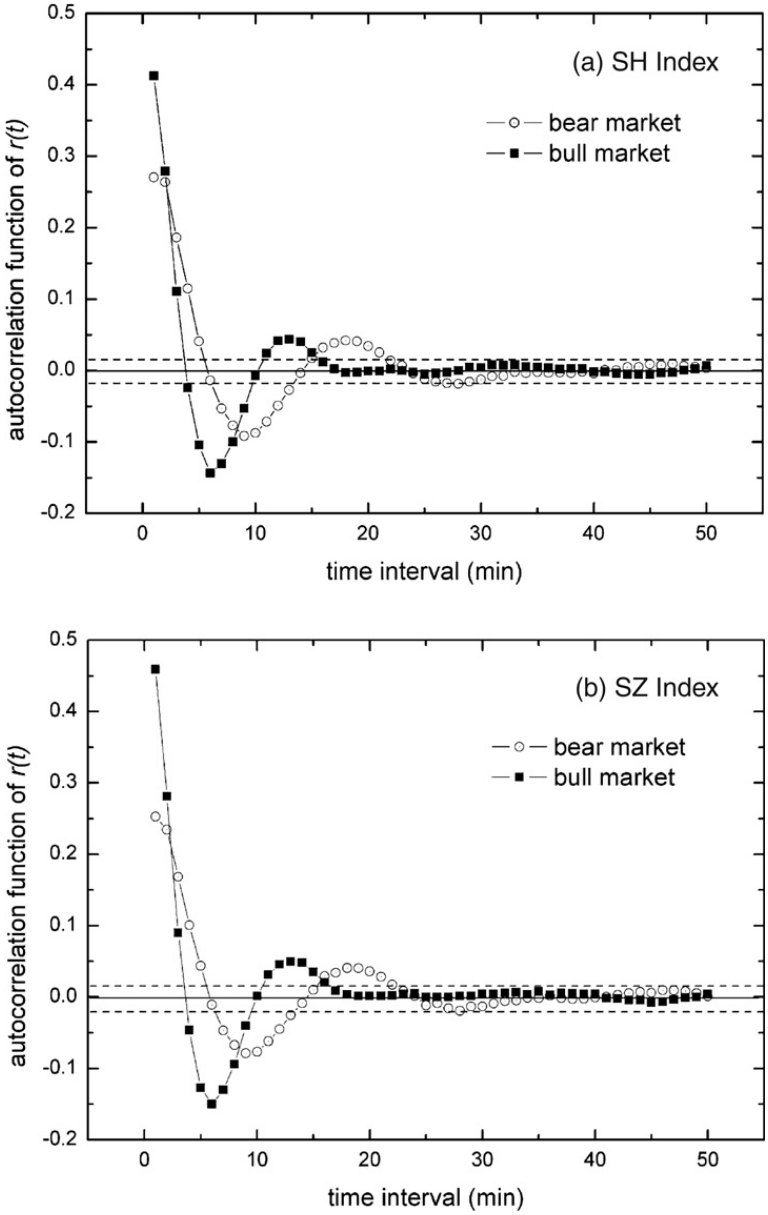

Fig. 3. The autocorrelation function of (a) SH Index and (b) SZ Index log-return $r(t)$ in the bear market (in circle) and the bull market (in square) respectively. We calculate the standard deviation σ_e for the autocorrelation function between 30 min and 50 min, and depict the $3\sigma_e$ as dashed lines. These plots show that the autocorrelation of the bull market has a stronger negative correlation and a shorter correlation time (after which time the autocorrelation function value cannot be distinguished from zero) than that of the bear market.

where $\langle x \rangle$ means the average value of x over the entire time series. Given that t' in $C(t')$ runs from $[1, T]$, then we divide the range $[1, T]$ uniformly into windows of size t, and fit $C(t')$ to a linear function $C_t(t')$ in each window. Eventually, we get the DFA function, which is defined as

$$F(t) = \sqrt{\frac{1}{T}\sum_{t'=1}^{T}[C(t') - C_t(t')]^2}. \tag{5.2}$$

Typically, $F(t)$ will increase with the window size t and obeys a power-law behavior, which indicates the presence of scaling (Fig. 5).

$$F(t) \sim t^{\theta}. \tag{5.3}$$

For completely uncorrelated series, such as white noise, $\theta = 0.5$. The condition $0.5 < \theta < 1$ indicates persistent long-range autocorrelation that a relatively large return is more likely to be followed by a large return and vice versa, while $0 < \theta < 0.5$ indicates an anti-correlation that the large and small values of time series are more likely to alternate. The case $\theta = 1$ corresponds to $1/f$ noise. If $\theta > 1$, the time series is considered to be unstable and will have a very smooth landscape (compared to that of white noise) of Brownian noise when $\theta = 1.5$.

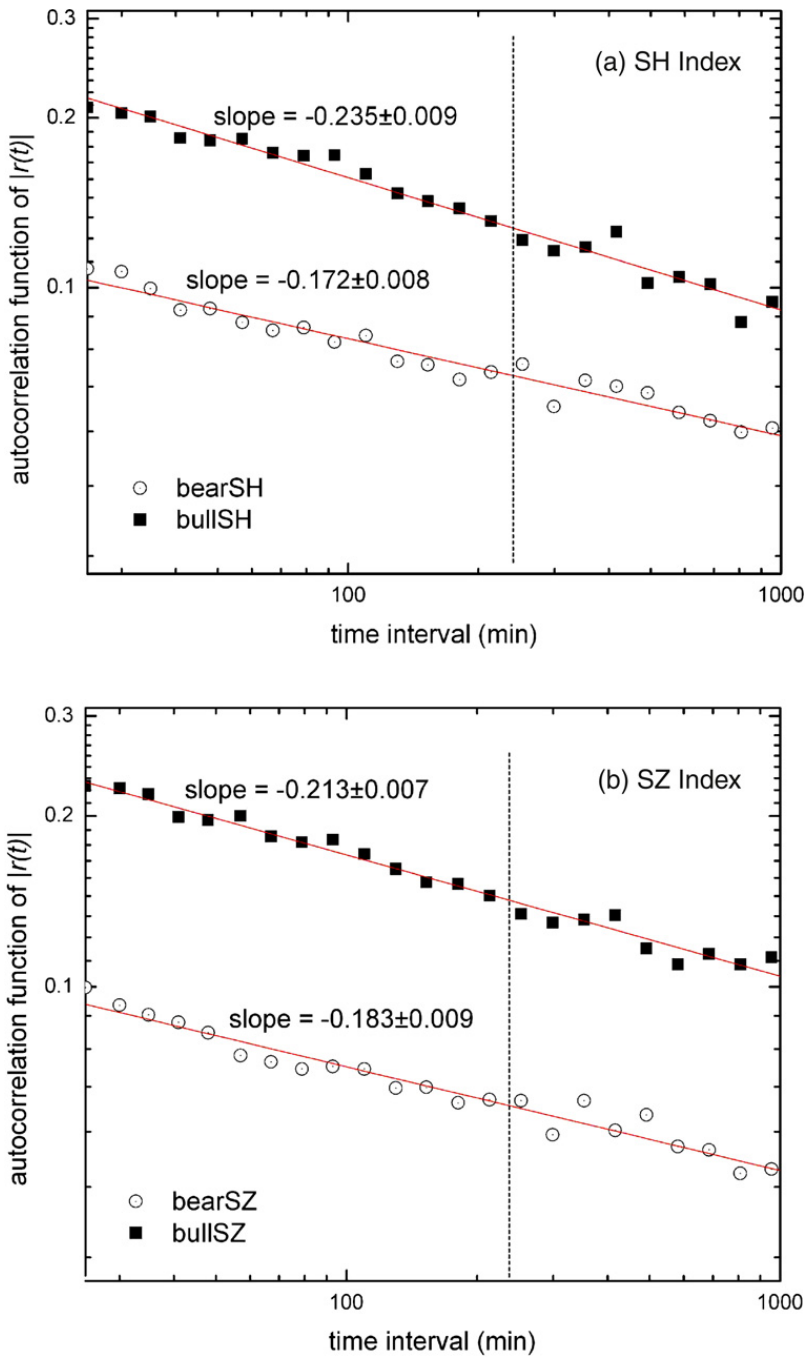

Fig. 4. The autocorrelation of the absolute log-return $|r(t)|$ of (a) SH Index and (b) SZ Index in the bear market (in circle) and the bull market (in square), in which the intraday pattern is removed. The bull market has a larger value of autocorrelation function for $|r(t)|$ than the bear market. The vertical dashed line indicates the time interval of one day (240 min).

In Fig. 5, we can see that the DFA functions of SH Index and SZ Index behaves similarly. In the bull market, the exponent θ approximates 0.80; while θ approximates 0.70 in the bear market. This implies that the long-term autocorrelation of the stock indices is stronger in the bull market than in the bear market, which is in accordance with our conclusion in Section 4.

6. Conclusion and discussion

We have found three statistical features of the indices' log-return $r(t)$: (i) the CDF curve of $r(t)$ in the bull market is similar to that in the bear market; (ii) the autocorrelation function of $r(t)$ in the bull market has a stronger negative correlation and a shorter correlation time than that in the bear market; (iii) the bull market shows stronger long-term correlation than the bear market.

Why does the autocorrelation of $r(t)$ have different behaviors in the two market phases? Bouchaud and Potters attributed it to the liquidity of the market in their book [5]. Here we depict the trading volume from Jan 2001 to Oct 2007 in Fig. 6. The

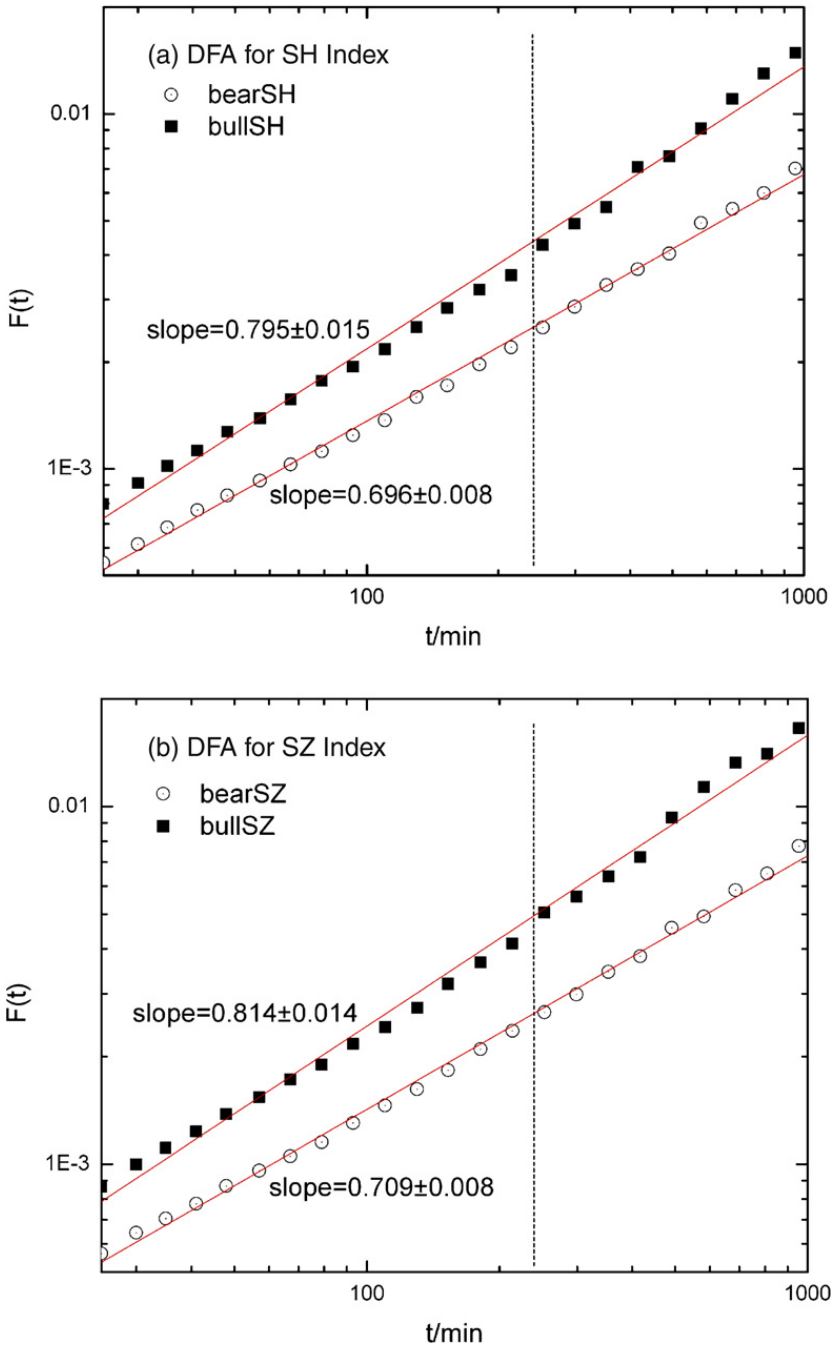

Fig. 5. (a) The DFA functions of SH Index on the log–log scale. (b) The DFA functions of SZ Index on the log–log scale. In both plots, the circles and squares indicate the bear period and the bull period respectively. The vertical dash line corresponds to 1 trading day (240 min).

huge volume after Jan 2006 shows that the market is more liquid in the bull market. So Bouchaud and Potters' conclusion, the less liquid the market, the longer the correlation time, is verified.

However, we want to investigate it further. Larger trading volume in the bull market is caused by increasing money and investors participating in the market. If we believe that there are a certain proportion of speculators in new investors, then there will be more speculators in the bull market, who take advantage of all kinds of arbitrage opportunities, including the autocorrelation pattern of $r(t)$. Some speculators will try to take action before others do or respond more rapidly to the trading signals than they do in the bear market, causing the correlation time in the bull market to become shorter. (But the correlation time cannot be eliminated due to transaction fees.) In a word, it is the increased speculators as well as their confident attitude to the trading signals that cause the shorter correlation time in the bull market.

But why does the autocorrelation of $|r(t)|$ have stronger long-term correlation instead of weaker correlation (analogous to the shorter correlation time of $r(t)$)? Again, we ascribe it to the special characteristics of the participants in the bull market. In the recent bull market in China, many people believed that the stocks' prices will rise "forever". Articles preaching that the golden ten years had come could be found in almost every newspaper. In such an irrational bull market, investors

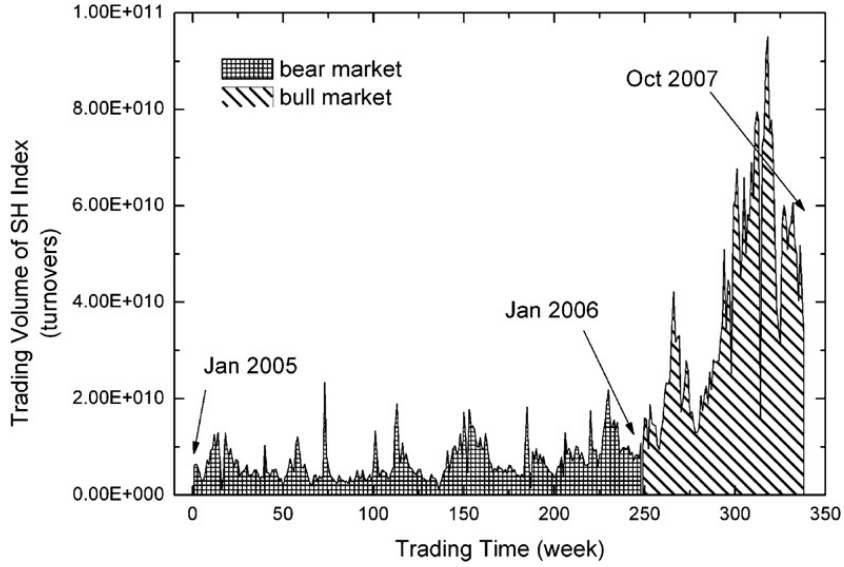

Fig. 6. The evolution of the trading volume for SH Index from Jan 2001 to Oct 2007. The two market phases are marked differently. The plot shows that the bull market is more liquid than the bear market.

were repeating such a pattern: they wait and watch the trading signals, and take part in the market in a hurry whenever they think there is a trading signal. So, their herding behavior includes some kind of intermittency, which could produce bursty structures in the trading volume. Meanwhile, the herding behavior will also induce some longer repercussions of the fluctuations in the return, which could enhance the long-term correlation of the fluctuation. Thus, it seems to be the combination of the clustering structure in the time series of the trading volumes and the longer repercussions of the fluctuation that causes the long-term correlation in the absolute return. In the bear market, their response becomes more sluggish, creating less bursty structures. However, it needs applying agent-based models to verify our assumptions, which is not covered in this article.

The behavior of the stock market is difficult to understand. This article is intended to study the stock market from a new perspective, helping to understand novel statistical properties in economic systems. However, more work should be done to confirm or improve the conclusions in this article.

Acknowledgements

This work was supported by the National Student's Innovation Training Program and by the National Fund for Talent Training in Basic Science (No J0730310). We thank Professor H. R. Ma of Shanghai Jiao Tong University for his helpful assistance. J.P.H. also thanks Professor Ping Sheng of the Hong Kong University of Science and Technology for his hospitality and the discussion conducted while visiting his group for two months in Summer 2008. The revised version of this work was just completed in the duration of the visit.

Appendix. Intraday pattern

The intraday pattern has been widely discussed [18,19], which is defined as

$$D(t'_{day}) = \frac{1}{N}\sum_{i=1}^{N}\left|r_i(t'_{day})\right|, \tag{A.1}$$

where N is the total number of trading days, and t'_{day} is the time in any trading day (Fig. 7 (a)). This intraday pattern will cause artificial period of correlation, so we remove this influence by

$$r'(t'_{day}) = \frac{r(t'_{day}) \times \langle |r(t)| \rangle}{D(t'_{day})}, \tag{A.2}$$

where $\langle x \rangle$ denotes the average value of x. Fig. 7 (b) shows the difference between $|r(t)|$ (with intraday pattern) and $\left|r'(t)\right|$ (without intraday pattern). After the intraday pattern is removed from the original time series, the periodic autocorrelation function is smoothed.

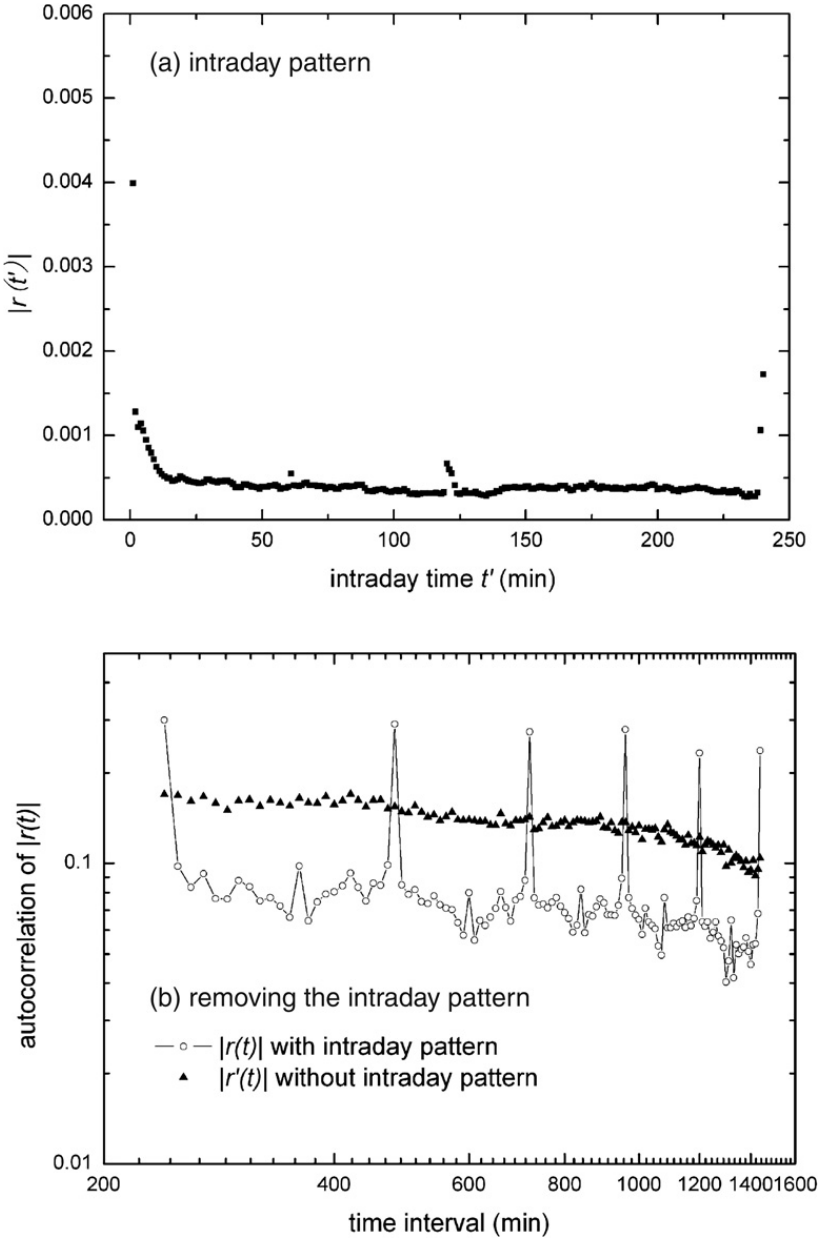

Fig. 7. (a) The 1-min interval intraday pattern for the index CSI-300 in the bull market, intraday $|r(t')|$ is the mean return over all trading days. (b) The autocorrelation function of $|r(t)|$ with intraday pattern (in circle) and without intraday pattern (in triangle). After the intraday pattern is removed from the original time series, the periodic autocorrelation function is smoothed.

References

[1] B. Mandelbrot, J. Business 36 (1963) 394.
[2] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics, Nature 376 (1995) (2000) 46. Cambridge University Press.
[3] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, Phys. Rev. E 60 (1999) 6519.
[4] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, H.E. Stanley, Phys. Rev. E 60 (1999) 5305.
[5] J.P. Bouchaud, M. Potters, Theory of Financial Risk, Cambridge University Press, 2000.
[6] P. Gopikrishnan, V. Plerou, Y. Liu, L.A.N. Amaral, X. Gabaix, H.E. Stanley, Physica A 287 (2000) 362.
[7] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.K. Peng, H.E. Stanley, Phys. Rev. E 60 (1999) 1390.
[8] T. Lux, Appl. Fin. Econ. 6 (1996) 463.
[9] V. Plerou, P. Gopikrishnan, X. Gabaix, H.E. Stanley, Phys. Rev. E 66 (2000) 027104.
[10] R. Storer, S.M. Gunner, Internet J. Modern Phys. C 13 (2002) 893.
[11] K. Matia, M. Pal, H. Salunkay, H.E. Stanley, Europhys. Lett. 66 (2004) 99.
[12] Z.F. Huang, Physica A 287 (2000) 405.
[13] C. Yan, J.W. Zhang, Y. Zhang, Y.N. Tang, Physica A 353 (2005) 425.
[14] J.W. Zhang, Y. Zhang, H. Kleinert, Physica A 377 (2007) 166.
[15] T. Qiu, B. Zheng, F. Ren, S. Trimper, Physica A 378 (2007) 387.
[16] C.K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Chaos 5 (1995) 82.
[17] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685.
[18] L. Harris, J. Fin. Econ. 16 (1986) 99.
[19] A. Admati, P. Pfleiderer, Rev. Financial Stud. 1 (1988) 3.