International Journal of Economics, Business and Management Research

Vol. 2, No. 01; 2018

ISSN: 2456-7760

RESEARCH ON THE FAST DETECTION METHOD OF V/S LONG MEMORY TEST FOR FINANCIAL TIME SERIES —EMPIRICAL STUDY ON THE FEATURE POINTS OF EXTREMUM

Wang Qian[1] ; Guan Heshan[2]

([1,2], the School of Economics and Law in University of South China, Hengyang Hunan Province in China,421001)

Abstract

To explore the inherent law of financial market price fluctuation and the characteristics of financial time series in the age of mass data, which is significant to understand and predict the trend of financial market. Nowadays, data becomes have more various types and large volume, under the background of massive data, we find that: the traditional long memory test method's computation speed cannot meet the requirement of modern long memory test research. The paper takes V/S method as an example, in order to compressing the sample size, reducing computational complexity and simplifying rapid calculations, it adopts the characteristic points of extremum. Empirical findings: the results of V/S long memory test are nearly unanimous between the fast detection method and the original one, especially for k value is 1, meanwhile, it could greatly improved time efficiency under guarantee the accuracy of this test.

Keywords: The Age of Big Data; Financial Time Series ; Long Memory; V/S Test

1. Introduction

With the advent of the big data era, the data types of Ultra High Frequency(UHF), High Frequency(HF) and the interval between days, weeks and months have emerged in large numbers. Meanwhile, the differences between acquisition data time and financial markets (China, Hong Kong,US etc.) all result in increased in the amount and variety of data stored, and it's directly present obviously a picture of big data character, namely: Volume, Velocity, Variety and Value characteristics[1]. This phenomenon makes the original traditional data storage and processing methods gradually unable to meet the needs of modern financial analysis and decision-making, so it's particularly urgent to find a fast method to test the characteristics of financial time series. On the one hand, it can retain the basic characteristics and trends of financial time series under the original massive data, on the other hand, it can compressed sample size and then achieve the purpose of simplification and fast the operation. At the same time, it can greatly reduce the situation of experimental interruption, waste of resources and other problems because of the huge amount of data. Therefore, it's particularly important to explore a rapid test method of mass financial time series in big data era.

In the literature research process, it was found that: due to the method based on the extreme feature extraction can be used to improve the accuracy and efficiency of subsequence matching with the original sequence in the practical application, and achieved a good results, so the feature points based on extreme points have been used widely in image, video, hydrology and the simulation model of system. However, due to the differences of discipline span, its application in financial time series similarity is not enough. The typical study is Wu Xueyan's[2], by identifying the extreme feature points in time series, the multilevel extreme value partition method is used to partition the long sequence according to the extreme points. It is proved that this method can greatly improve the efficiency of similarity search process under the condition of ensuring accuracy.

The main purpose of this paper aims to study a quick test method of long memory time series by analyzing financial time series data, and plans to use time series data of different financial markets, different period and different types to test it in the follow-up study. And this will enrich the test method of long memory time series in theory. Maybe, it can also conduct a systematic empirical verification with long memory time series data in practice under the background of big data, and provide a reference for the financial investment.

2. Summary of long-term memory

Long-term memory is also termed the long-range correlation or persistence. It describes the long-term correlation of sequences, which were performed as the self correlation coefficient of time series decreased slowly[3]. But after a long lag order, there is a significant correlation still in the sequence of contiguous variables, and the relationship will last for a long time. Long memory is used to describe the dependence relationship between two nonadjacent variables and makes it clear that the time series is acyclic. If the dynamic dependence structure show some characteristics of long time dependence,they will make it possible that predict the future trend by using historical information in time series. This can also help us grasp the characteristics of our financial time series accurately, meanwhile, understand and predict the trend of the financial market, also provides the possibility of establishing the forecasting model which can meet it's characteristic. The memory of the sequences be seemed as one of the key factors to establish prediction model, research on financial asset price and the essence of memory under the big data era can not only provide basis for the macro-control policy formulation and decision making, but also offer a practical reference for the organization and individual investors.

In the last ten years, the research on long memory had achieved fruitful results, but the conclusion is different for the long memory of the Chinese stock market. He Xingqiang and Li Zhongfei (2006) made a V/S test on China's stock market(in Shanghai and Shenzhen) and the result shows that the stock market in China is anti-persistence rather than long-term memory[4]. Xu Youzhuan (2010) found that the R/S test was more likely to accept long memory, and drawn the conclusion that the empirical results obtained by using the R/S method and MR/S method

ISSN: 2456-7760

should be interpreted with caution[5]. Li Zhisheng et al. (2011) found that the stock market in China has a significant long-term memory and has gradually increased in recent years by using sequence correlation test and R/S test[6].Yuan Ying et al. (2012) also discovered that the Chinese stock market returns and turnover series have long memory characteristics by using R- S method, MR/S method, KPSS test and Granger causality tes t[7].Xiao Weilin et al. (2013) studied the impact of long memory modeling, and found that the pricing results of different models all proved that financial assets have long-term memory[8].Chen Langnan et al. (2013) constructed an self-adaptive asymmetric HAR-CJ-D- FIGARCH model. It is found that the high frequency volatility of Chinese stock market is characterized by long memory, structural mutation, asymmetry and day of the week effect[9].Dai Yingjie et al. (2013) point out that the fluctuation of commodity housing price in China have identified with long memory, aggregation effect and heteroscedasticity by using FIGARCH model[10].

In view of there are many methods had used and many previous researches had made on the long memory of time series, this paper plans take V/S method as an example to analyse the test results and efficiency between the original V/S method and the fast detection method of improved V/S test.

3. V/S method and demonstration

V/S method

The discriminant of the V / S method is as follows:

$$(V/S)_n = \frac{1}{nS_n^2}\left[\sum_{k=1}^{n}\left(\sum_{t=1}^{k}(x_t-\bar{x})\right)^2-\frac{1}{n}\left(\sum_{k=1}^{n}\sum_{t=1}^{k}(x_t-\bar{x})\right)^2\right]; \quad (1)$$

$$S_n=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2};\ \bar{x}=\frac{1}{n}\sum_{t=1}^{n}x_t;$$

V/S method’s calculation of H values is in accordance with R/S method, they were both calculated their slope with making the line regression of log(n) ~ log(V / S) by using least squares method, but the difference is H value of V/S method is only half of the slope. In addition, the value of H regards as the criteria of the long memory of sequence, the judging scope also coincides with the R/S method[11]. There are some inconvenient problems in testing can be found: When the sample size is great(such as>1000), the calculation speed and the effect is not so positive because of the difference structural mechanism(the internal circulation of deviation from the mean is N squared rather than N). Especially in massive data processing, this problem will be more prominent and even lead to experimental interruption or operation

ISSN: 2456-7760

paralyzed problem in software system with limited memory. Just because the computing speed of the V/S method under the massive data is far from meeting the needs of modern research, this paper expects to use feature point extraction method of extremum to achieve the purpose of compressing sample size, reducing complexity and simplifying computing.

The method of feature point extraction of extremum

Let a time series of length n be the equal of $Y = \{y_1, y_2,..., y_n\}$, i indicates the ith element of the current calculation position of the y_i series, k represents the currently calculated neighborhood (i.e., the front and back k items which centered with y_i)[11]. It is generally considered that if the y_i value of the neighborhood in which the radius is k, and y_i is larger than all the values in the range of the neighborhood, it can be considered as the maximum value in the range of this domain. Accordingly, if the value is small than all sequence values in the neighborhood range, it is considered to be the minimum value in this domain.It can be seen that the range of neighborhood (that is, the value of the k) can directly affects the sample size of the time series after extracting the extremum.

Demonstration

Data preprocess and collection

With the help of data processing software Matlab, the paper tried to use the V/S method to make a long memory test with the Shanghai Stock Exchange 50 Index (SSE50 for short. This article uses codes 1-50 to represent each Shanghai Stock Exchange 50 Index, the detailed stock codes are shown in the appendix) which during January 1st 2004 to February 28th 2014. At the same time, the paper adopts the method of feature point extraction of extremum to compress the original series data and makes them have a the long memory test by V/S way, and then the calculation results and efficiency of the two kinds of data are compared and analyzed. Thus the evaluation was finished of the fast algorithm based on feature point extraction of extremum.

Generally, logarithmic difference method is often used to preprocess the financial time series

of stock index, which can reduce the correlation of time series and the influence of trend on the test results. If the original time series is $\{x_i, i = 1,2,..., n\}$, after using the logarithmic difference method, the stock index yield can be expressed as:

$$R_i = \ln(x_{i+1}) - \ln(x_i) \qquad (2)$$

xi represents the closing price of the i th day, and xi□1 represents the closing price of the（i +1）th day,

Analysis result of the V/S Test of yield Series.

Table 1 Empirical results of daily yield of 50 components of SSE50 by V/S method

ISSN: 2456-7760

Ordinal	H value	Ordinal	H value	Ordinal	H value	Ordinal	H value	Ordinal	H value
1	0.7108	11	0.5831	21	0.4411	31	0.5285	41	0.3761
2	0.4768	12	0.621	22	0.6528	32	0.51	42	0.5203
3	0.5888	13	0.5823	23	0.3883	33	0.5671	43	0.356
4	0.5447	14	0.5975	24	0.5617	34	0.3303	44	0.5966
5	0.2896	15	0.3757	25	0.5894	35	0.6652	45	0.6109
6	0.7265	16	0.6498	26	0.5373	36	0.5211	46	0.3798
7	0.6675	17	0.4938	27	0.4784	37	0.4846	47	0.7363
8	0.5485	18	0.466	28	0.4419	38	0.5676	48	0.5382
9	0.6765	19	0.7026	29	0.6887	39	0.4757	49	0.5754
10	0.5217	20	0.321	30	0.5902	40	0.6672	50	0.6830

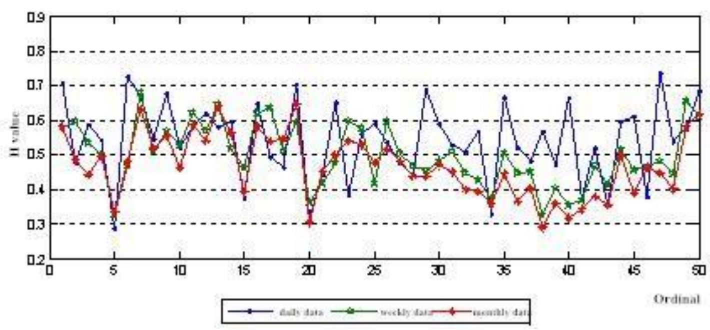

Fig.1 the distribution of H values of the V / S test for the daily, weekly and monthly yield series of the 50 component stocks of the Shanghai Stock Exchange

From table1, it is not difficult to find that most of the components in the SSE50 tend not to reject the original hypothesis of having long memory. Based on the analysis of the Fig.1, it is found that from the value of H value, H value of daily data is generally larger than that of other types of data, and more inclined to have long memory; The data of week and month are more likely to have the characteristics of reverse memory. Among which, the number of long memory series in different data sampling methods (day, week and month) is 34, 25 and 18 respectively. The number of sequences that met the criteria of long memory showed a decreasing trend. Meanwhile, there are about ten stocks whose H value is close to the weak memory of 0.5, which may be related to the principle that the rate of return is calculated by nonlinear transformation. The non-stationary data feature is eliminated by this transformation, which makes the result of long memory test become more objective.

Time consuming

ISSN: 2456-7760

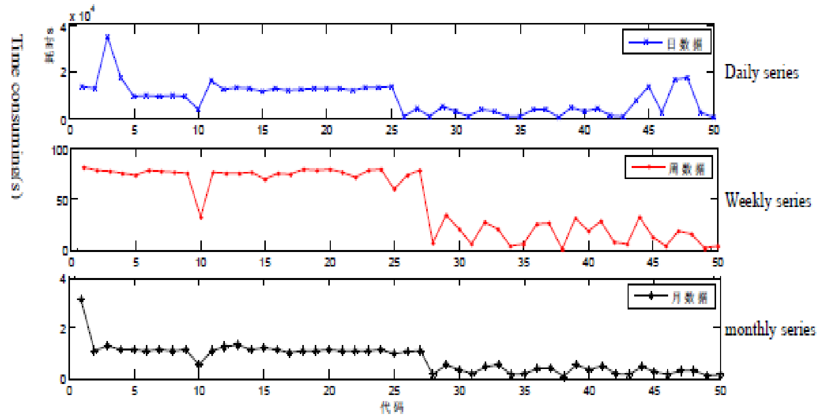

Fig.2 the time consuming of the V/S test for the daily, weekly and monthly yield series of the 50 component stocks of the SSE50

Although the V/S method is widely used in long memory testing of time series. Giraitis et al. replaced the polar difference of cumulative deviation of samples in the R/S test with sequence variance. Also three methods have been studied by Giraitis from the theory and Monte Carlo simulation: the results of MR/S test, KPSS and V/S statistics, and he found that V/S analysis is more efficient and robust than others in long memory test[11]. It is precisely because the V/S method uses the sequence variance rather than range operation that the computational complexity and frequency of this method is greatly increased to n square. When the sample size is become large, the calculation rate will be obvious slow. And Figure2 shows that there is a significant positive correlation between the amount of time and data. The average duration of the daily, weekly and monthly yield series is 8375.21 seconds, 46.59 seconds and 0.77 seconds respectively. Also the time taken is significantly under different sampling patterns, as shown in Table 2.

Table 2 the max and min time consuming series of the V / S test for the yield of component stocks

Statistics	Daily sequence			Weekly sequence			Monthly sequence		
	H value	Time consuming	Amount of data	H value	Time consuming	Amount of data	H value	Time consuming	Amount of data
Max	0.7363	18228.3019	2421	0.6815	151.9597	515	0.6453	1.1946	122
Min	0.2896	61.4849	516	0.3252	0.8409	112	0.2925	0.0177	27

Analysis of V/S Test results based on feature point extraction of extremum

Because the calculation speed and efficiency of V/S test is not ideal in the case of large amount of data, the method of extracting extreme feature points is considered in this paper to achieve the purpose of compressing the sample size and accelerating the calculation.In view of the abundance of daily return rate series data, it is convenient to select neighborhood and divide the fields, so the paper takes the data of daily rate of return as a sample to analyze. This paper compresses the sample firstly by the previous method, then preprocesses the compressed data by formula (2), and finally, makes the V/S long memory test of the processed data by formula (1). At the same time, the results and consuming time of the whole operation and the data sample size of each component are reported as follows.

On account of the difference of k value (that is, the difference of neighborhood) will directly affect the size of subsample size after the original sample is compressed, the article have made an empirical study on k when the k is taken at 1 , 3 , 5 and 7 respectively. At the same time, emphasis is placed on the effectiveness and reliability which compared the V/S test under the fast algorithm of extremum feature point with the original method. From Fig.3 can be found that: when k takes different values, the variation trend of H value based on extremum point is basically the same under V/S method. However, because the value of k will directly affect the amount of data in the return sequence, there is still a big difference in the computational efficiency (especially in time consuming).

From Table 3, we can see that the value of k has a direct and critical effect on the result of the operation. The results of H value based on extreme point algorithm under different k values are basically inconsistent. In generally, with the increase of k value, H value has a decreasing trend, and the time consuming is also reduced. This shows that if the speed of the original sequence operation is faster, the result of H value obtained may deviate further from the result of the original method. But the smaller the value of k and the smaller the selected neighborhood, then the value of H will be as close as possible to the original value. Taking k =1 as an example, the average calculation speed is shortened to about 1/4, and the calculation rate is greatly improved. Compared with the time consuming of taking a larger value of k (such as k =7), it still has advantages in approaching the real H value, although the time consuming of calculation is more.

Table 3 V/S test's empirical results of daily yield series based on feature point extraction of extremum

Ordinal	H value				Data Size	Ordinal	H value				Data Size
	k=1	k=3	k=5	k=7			k=1	k=3	k=5	k=7	
1	0.7046	0.6569	0.6236	0.6113	2390	26	0.5260	0.4547	0.4507	0.4680	2351
2	0.4606	0.3741	0.3585	0.3308	2359	27	0.4754	0.4440	0.4276	0.3818	2389
3	0.5797	0.5419	0.4729	0.4580	2364	28	0.4205	0.4000	0.3563	0.3593	1030

ISSN: 2456-7760

4	0.5237	0.4638	0.4622	0.458	2409	29	0.6616	0.6303	0.5959	0.5575	1817
5	0.2867	0.2461	0.2841	0.2544	2380	30	0.6037	0.5370	0.4901	0.4819	1548
6	0.7179	0.6425	0.5850	0.6022	2416	31	0.5133	0.3649	0.3169	0.3233	997
7	0.6522	0.6562	0.6185	0.5943	2384	32	0.5064	0.4408	0.4086	0.3932	1694
8	0.5323	0.5199	0.4767	0.4639	2413	33	0.5508	0.4919	0.4658	0.4652	1546
9	0.6636	0.6285	0.5918	0.6157	2316	34	0.3074	0.2877	0.2739	0.2341	872
10	0.5205	0.4135	0.3953	0.3863	1822	35	0.6386	0.5094	0.5095	0.5067	991
11	0.5778	0.5403	0.5259	0.5071	2419	36	0.5121	0.4106	0.4180	0.3979	1635
12	0.6157	0.5744	0.5491	0.5562	2375	37	0.4584	0.3920	0.3730	0.3347	1636
13	0.5857	0.5049	0.4790	0.4867	2421	38	0.5573	0.3784	0.3193	0.2886	527
14	0.5833	0.5069	0.5044	0.4544	2387	39	0.4617	0.4260	0.4228	0.4132	1760
15	0.3688	0.3082	0.2767	0.2804	2306	40	0.6512	0.5822	0.5526	0.5112	1490
16	0.6313	0.5756	0.5303	0.5063	2394	41	0.3549	0.3044	0.3336	0.3301	1723
17	0.4889	0.3901	0.4372	0.4157	2334	42	0.5105	0.4438	0.5559	0.4120	1103
18	0.4456	0.4018	0.3699	0.3308	2362	43	0.2789	0.2346	0.2150	0.1907	966
19	0.7053	0.6693	0.6228	0.5656	2396	44	0.5575	0.4709	0.4430	0.4111	1788
20	0.3183	0.2748	0.2614	0.2478	2400	45	0.5917	0.5187	0.4964	0.4295	1327
21	0.4379	0.3883	0.3772	0.3508	2393	46	0.3470	0.3368	0.3511	0.3020	844
22	0.6420	0.5848	0.5647	0.5338	2337	47	0.7133	0.6875	0.6222	0.5748	1525
23	0.3630	0.3177	0.2829	0.3016	2410	48	0.5187	0.3899	0.3340	0.3142	1404
24	0.5572	0.5657	0.5477	0.5477	2416	49	0.5499	0.4231	0.3875	0.3911	516
25	0.5792	0.5095	0.4883	0.4516	2212	50	0.6561	0.6588	0.6008	0.6014	861

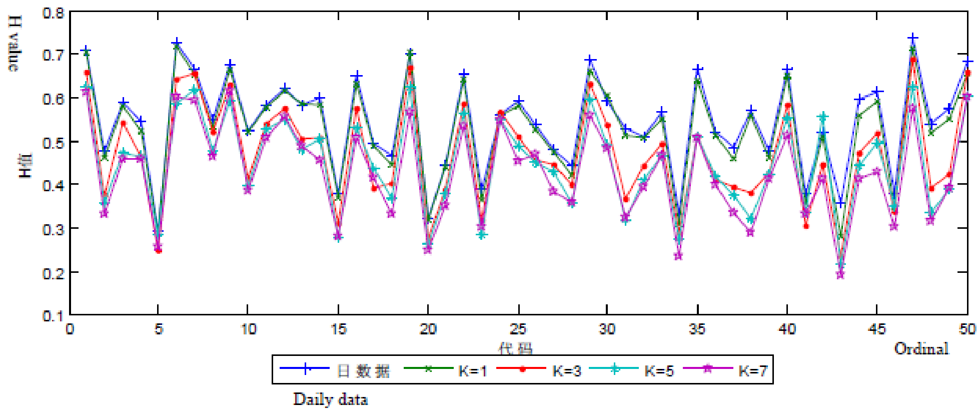

Fig.3 the distribution of H values in different k numeric value

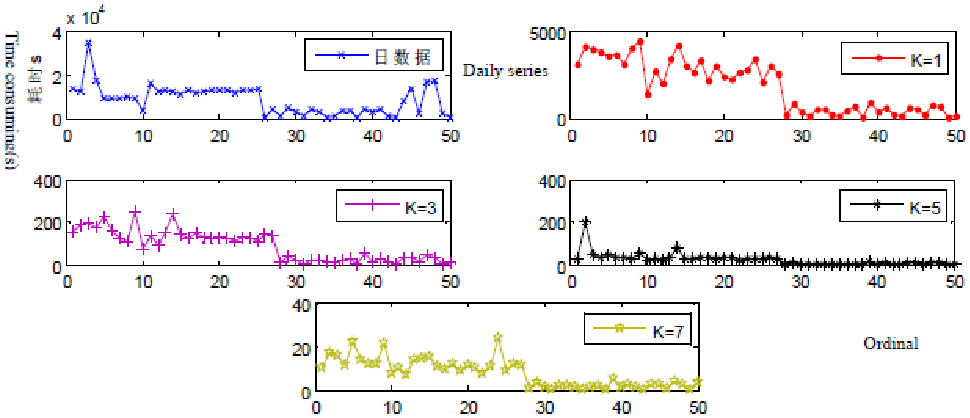

Fig.4 time consuming of V/S test for different k numeric value

And interestingly, it is not difficult to find a certain rule in the calculation results under different k values. The calculation time of k = 1 is nearly 20 times of that k = 3, and that of k = 3 is nearly four times as long as that of k = 5, and the time consuming of k = 5 is nearly twice as long as that of k =7. This shows that there is a rule for calculating H value under different k values. If we want the calculation results to be close to the reality and the calculation speed is faster, it is more feasible to adopt the smaller k value. Meanwhile, we still need to combine the actual calculation requirements.

4 . Conclusions

In this paper, the yield data of SSE50 index and its constituent stock in the past ten years are tested by the long memory test based on feature point extraction of extremum and its H index is obtained, and something can be found on the basis of reference to the existing discriminant rules: the traditional long memory test method can not meet the needs of modern long memory test research. The V/S method is taken as an example in this paper, through the method of feature point extraction of extremum to achieve a goal of compressing sample size and then achieve the purpose of simplification and fast the operation. It is found that the accuracy and reliability of the proposed method based on extreme feature extraction are similar with the original algorithm (especially when k = 1), and the same time, this method can greatly improve the time efficiency of the long memory test process without losing the accuracy.

References

[1] Meng Xiaofeng, Ci Xiang.Big Data Management:Concept, Techniques And Challenges[J]. Journal Of Computer Research And Development, 2013, 50(1): 146-169.

[2] Wu Xue-Yan, Huang Dao-Ping, Mo Zan.Time Series Similarity Matching Algorithm Based On Extreme Points[J].Application Research of Computers, 2010, 27(6).

[3] Qian Wang,Chun Liu, Heshan Guan, Zhichao Luo.Study On Long-Term Memory Characteristics For a Stock Market[J].Finance And Accounting Monthly,2015,17:126-129.

[4] He Xing-Qiang, Li Zhong-Fei.Long-Termmemory In Stock Returns Of Shanghai Stock Exchange:Evidence From v/s Statistic[J].Systems Engineering-Theory & Practice,2006,12(12),47-54.

[5] Xu Youchuan.Study On The r/s Test Validity Of Long Memory Of The Financial Time Series[J].Chinese Journal Of Management Science,2010(8):204-209.

[6] Yuan Ying, Du Le-Lu, Zhuang Xin-Tian.An Empirical Research On Long-Term Memory Of Returns And Trading Volumes Of Stock Markets In China[J].Journal Of Northeastern University(Natural Science),2012(7):1056-1060.

[7] Li Zhi-Sheng, Liu Zheng-Jie. Short And Long Memory In Return On Assets: A Dynamic Analysis Of The Efficiency Of China’s Stock Market[J].Journal Of Jiangxi University Of Finance And Economics,2011(1):5-12.

[8] Xiao Weilin, Zhang Weiguo, Xu Weidong. Pricing Equity Warrants In a Bifractional Brownian Motion[J].Journal Of Systems Engineering ,2013,28(3):348-354.

[9] Chen Lang-Nan, Yang Ke.High-Friquency Volatility Features,Forcast Model And Performance Evaluation[J].Systems Engineering-Theory & Practice,2013,33(2):296-307.

[10] Daiying-Jie, Xie Yan, Feng Zhao-Zhen. An Empirical Analysis Of Long-Term Memory

Fluctuation Rate Of Housing Price[J].Journal Of Northeastern University(Social Science),2013,15(2):128-133.

[11] Giraitis l, Kokoszka p, Leipus r, Et Al. Rescaled Variance And Related Tests For Long Memory In Volatility And Levels[J]. Journal Of Econometrics, 2003, 112(2): 265-294.

Vol. 2, No. 01; 2018

ISSN: 2456-7760

Appendix1: the constituent stock code of SSE50 published in February 2014

Table 4: the stock code and number of SSE50

Ordinal	The Stock Code	Ordinal	The Stock Code	Ordinal	The Stock Code	Ordinal	The Stock Code
1	600000	14	600256	28	600999	42	601668
2	600010	15	600332	29	601006	43	601688
3	600015	16	600362	30	601088	44	601699
4	600016	18	600406	31	601117	45	601766
5	600018	19	600489	32	601166	46	601818
6	600028	20	600518	33	601169	47	601857
7	600030	21	600519	35	601299	48	601899
8	600031	22	600547	36	601318	49	601901
9	600036	23	600549	37	601328	50	601989
10	600048	24	600585	38	601336	—	—
11	600050	25	600637	39	601398	—	—
12	600104	26	600837	40	601601	—	—
13	600111	27	600887	41	601628	—	—