Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 22 (2013) 1172 – 1181

Procedia Computer Science

17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013

Predicting the security levels of stock investment by using the RMT-test

Xin Yang[a] , Yuuta Mikamori[a], Mieko Tanaka-Yamawaki[a,] *

[a]Information and electronics Department, Graduate School of Engineering, Tottori University, Tottori 680-0945, Japan

Abstract

The authors propose to use the degree of randomness of high frequency price time series for the purpose of measuring the security levels of stock investments. The RMT-test is employed as a tool to measure the randomness. The data to be analyzed are the tick-wise price time series of selected stocks in the Tokyo Stock Exchange Market for three years from 2007 to 2009. The result shows that the stock of the highest randomness is a stable stock that belongs to the sector of electric/gas power supply, which turns out to be more profitable than the Nikkei Average Price throughout the following year. This indicates that the suitable stocks to invest under a bear market have higher randomness that belongs to the category of 'defensive' stocks, according to the new classification method introduced by Tanaka-Yamawaki, et. al., while the suitable stocks to invest under a bull market have lower randomness that belong to the category of 'outer demand' and 'market sensitive' stocks in the same classification method.

Selection and peer-review under responsibility of KES International

Keywords: Randomness, RMT-test, RMT-PCA, tick-wise stock price;

1. Introduction

By the rapid progress of digitization in recent years, acquisition and storage of huge amount of data has become available in various fields such as weather, health, finance and census. This situation requires development of effective technology, in order to analyze such heavy data for various purposes. The Principal Component Analysis based on the Random Matrix Theory (RMT-PCA) has been proposed as a technique to separate the random components and the correlated components, in order to extract useful information out of massive amount of data. The authors have proposed to measure the randomness of a given long time series called RMT-test [1], and have demonstrated the effectiveness of the RMT-test by measuring the randomness of

* Corresponding author. Tel.: +81-0857-31-5223; fax: +81-0857-31-0879.
E-mail address: mieko@ike.tottori-u.ac.jp.

1877-0509
Selection and peer-review under responsibility of KES International
doi:10.1016/j.procs.2013.09.204

the physical numbers and pseudo-random numbers [2]. However, the real advantage of the RMT-test is its applicability on the real-world data whose randomness is not very high. In this paper, we apply the RMT-test on tick-wise stock prices and attempt to use the randomness of price fluctuations as a new indicator for stock investments.

2. Review of RMT-test

The method used in this paper is to compare the eigenvalue distribution of the correlation matrix, between *N* pieces of length *L*, to the corresponding theoretical formula of the eigenvalue distribution derived from the random matrix theory in the limit of *N* and *L* going to infinity, keeping $Q = L/N$ as a constant. This method is applied to the stock market in 2002 by Plerou [3]. The outline of the method is as below.

A data sequence is cut into *N* pieces of equal length *L*, then shape them in an $L\times N$ matrix, by placing the first *L* elements in the 1st row of the matrix, and the next *L* elements in the 2nd row, and so on, by discarding the remainder if the length of the sequence is not divisible by *L*. Then each column of the matrix is normalized to have zero mean and single variance. By multiplying this matrix with its own transverse matrix, the correlation matrix *C* is constructed, which is a symmetric $N\times N$ matrix whose (i, j) element is the inner products between the *i*-th and the *j*-th columns of the $L\times N$ matrix. All of the *N* eigenvalues of the correlation matrix *C* is obtained by numerical calculation. The randomness of the sequence is measured by comparing the eigenvalue distribution to the corresponding theoretical formula, called Marcenko-Pastur distribution:

$$P_{RMT}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (1)$$

$$\lambda_\pm = (1 \pm Q^{-1/2})^2 \quad (2)$$

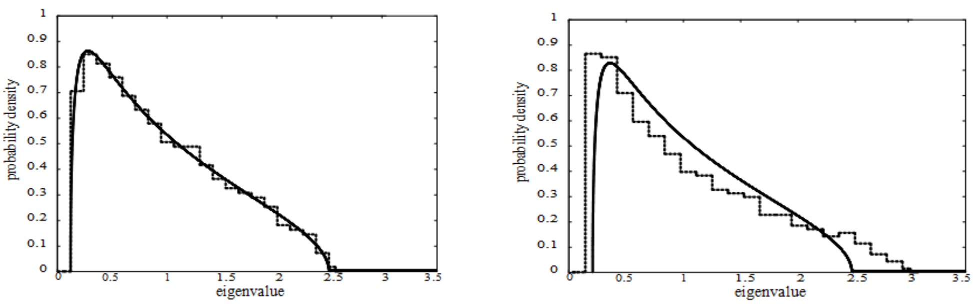

Fig.1. Example of the Qualitative evaluation of RMT-test (left: passed, right: failed)

The quantitative evaluation based on the moment method [4] compares the k-th moment of the obtained eigenvalues

$$m_k = \frac{1}{N}\sum_{i=1}^{N}\lambda_i^k \quad (3)$$

with the corresponding theoretical formula obtained from P_{RMT}

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+}\lambda^k P_{RMT}(\lambda)d\lambda \quad (4)$$

The difference between m_k and μ_k represents the degree of randomness of the data sequence. The authors have chosen to use the 6th moment (i.e., k=6) in order to define the level of randomness in the RMT-test [4], in which |Error| $= |m_6 - \mu_6|$ is employed to quantify the degree of randomness.

3. Randomness of stock prices

The qualitative evaluation of RMT-test employ a visual comparison of the eigenvalue distribution and its theoretical counterpart given by Eqs.(1) and (2). The left figure of Fig 2 is an example of low randomness given by the tick-wise price of the stock of Kao Corporation(code umber 4452), showing a clear protruding from the theoretical curve, in which |Error| = 0.199. The right part of Fig 2 is the example of the sequence made by a pseudo random number generator, the Linear Congruential Generator (LCG)[5], in which |Error| = 0.016. Namely, the |Error| of the real data is more than 10 times larger than the |Error| of the pseudo random numbers.

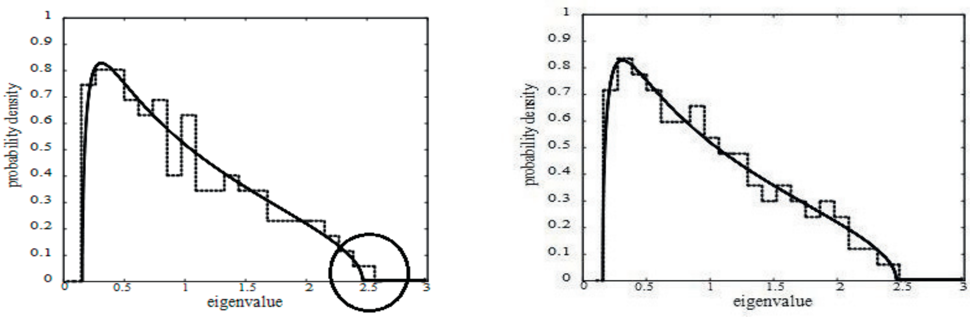

Fig.2. Two examples of randomness:(left)low randomness shown by the tickwise stock prices of Kao Corporation(Code 4452) and (right)high randomness shown by the LCG

3.1. Data processing

Tick data of stocks in TOPIX 500 from 2007 to 2009 per minute satisfying the experimental conditions are selected and used for analysis. Tick data mean the time series stamped in seconds or minutes which record the information of traded or quoted prices. Since trades or quotes may not occur at every time period, the lengths of the tick data are not fixed. For this reason, some work is required to prepare the fixed-length time series to serve for analysis. This process consists of the following 3 steps.

time	Stock			
	ticker1	ticker2	ticker3	ticker4
09:00	502	3810	1902	
09:01	502	3801	1906	
09:02	508	3806	1905	
...				
11:00	521	3788	1910	Exclude
12:30	515	3788		963
12:31	518	3792		953
...				
15:00	522	3820		853

Fig.3. Substitution of tick data for calculating the equal time correlation

(1) The blanks are filled by copying the previous data as long as the added part is less than the 20 percent of the total length as shown in Fig 3, in order to calculate equal time correlation of each stock price. For example, the ticker 3 and the ticker 4 in Fig 3 are both discarded. As a result, the data length (*L*) and the number of stocks (*N*) are different at each year. The values of *L* and *N* for each year used in this paper are summarized in Table 1.

Table 1. Data used for analysis

year	Data length (L)	Number of stocks (N)
2007	66338	211
2008	66338	240
2009	65945	229

(2) The price time series $p_1, p_2, ..., p_T$ are converted to the log-return time series $r_1, r_2, ..., r_T$ by means of

$$r_i = \log(p_i / p_{i-1}) \quad (5)$$

as usually done in the financial analysis for the sake of eliminating the unit/size dependence of different stock prices.

(3) The data sequence is cut into *N* pieces of equal length *L* by discarding the remainder when the length of the sequence is not divisible by *L* to construct a correlation matrix and obtain its eigenvalues and eigenvectors.

3.2. Choice of Parameter Q=4

The RMT-test has a free parameter, *Q=L/N*, to be determined beforehand. The condition required by the RMT is $Q>1$ and $N\rightarrow\infty$[6].

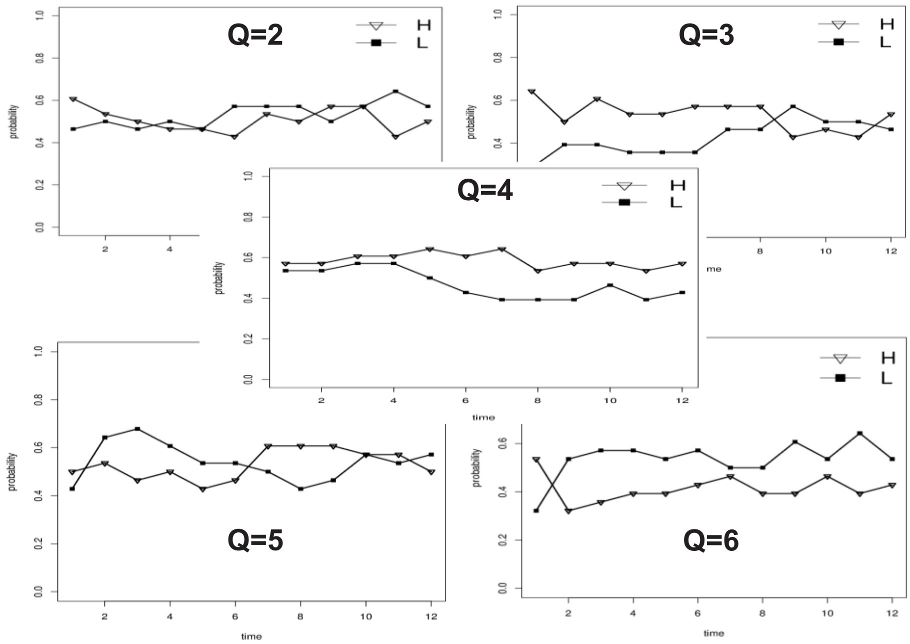

Fig.4. Optimal Q is 4 (H: highest randomness, L: lowest randomness)

It is known that $N > 100$ is necessary based on the numerical experiments using pseudo random generators. According to Table 1, the data length in 2009 is 65945= $N \times L$. Suppose $N=100$ is chosen, then $L=659$ thus $Q=6.59$. For $N > 100$, $L < 659$ and $Q < 6.59$. This means Q can be chosen in the range of $1<Q<6.59$. Therefore, $Q=2,3,4,5,6$ are the possible integer values for the parameter Q, and the analysis are done for each value of Q and the performance in the following year (January - December, 2010) of the stocks of the highest randomness (H) and the lowest randomness (L) are compared in Fig 4. Only the case of $Q=4$ exhibits the superiority of H to L throughout the year of 2010. The other values of Q do not show any clear sign of superiority of H.

3.3. Historical background

An overview of the economic situation of 2007-2009 is in order before getting into the analysis.

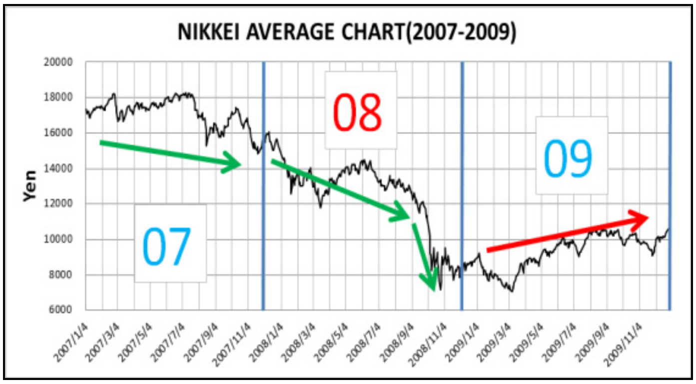

Fig.5. NIKKEI average chart from 2007 to 2009

Although the sub-prime loan problem occurred in 2007 and the stock market began to fall down, no serious collapse of companies occurred in Japan. However, the year of 2008 is quite abnormal, since the NIKKEI average fell down substantially due to Lehman shock, and 6 out of the top 10 record declines in day by day TOPIX before 2010 occurred in the year of 2008, as shown in Table 2.

Table 2. Top 10 recording of decline in TOPIX (before 2010)

rank	%	TOPIX	y/m/d
1	-14.2	1793.90	1987/10/20
2	-9.52	864.52	2008/10/16
3	-8.75	32.32	1953/3/5
4	-8.04	899.01	2008/10/8
5	-7.52	806.11	2008/10/24
6	-7.47	159.33	1970/4/30
7	-7.40	746.46	2008/10/27
8	-7.10	2069.33	1990/4/2
9	-7.10	840.86	2008/10/10
10	-7.05	889.23	2008/10/22

3.4. Experiment

The experiment processing is as follows:
⑴Compute the randomness of stocks by the RMT-test
⑵Sort the stocks according to the randomness to choose the first one and the last one
⑶Investigate the profit (log-return) of each stocks in the next year
⑷Analyze the relationship between the profit and the randomness of stocks

3.5. Result

⑴ Relation between the randomness and the stability of stocks

The ranking of randomness in 2007 is shown in Table 3, in which the stock of the highest randomness in 2007 is 9504 in the sector of Electric/Gas. The stock of the lowest randomness in 2007 is 7201 in the sector of Transportation and Equipment. Fig 6 shows the log-return of the top five stocks of code number 9504, 6460,9506,9508,4676 having the highest randomness comparing with the stock of the lowest randomness 7201. As a result, the top five stocks perform better than 7201 and effected very little by the financial crisis. On the contrary, the 7201 which has the lowest randomness in 2007 fell down enormously due to the effect of Lehman shock that occurred in September 2008. Based on this observation, the authors consider that the stock which has the highest randomness is stable and safe under a bear market.

Table 3. The ranking of randomness by using the tick data of 2007

Rank	Sector	Code	|Error|
1	Electric/ Gas	9504	26.4
2	Machinery	6460	37.6
3	Electric/ Gas	9506	38.2
4	Electric/ Gas	9508	43.3
5	Information & Communication	4676	44.9
	…		
207	Electric Appliances	6506	740.9
208	Nonferrous Metals	5802	797.3
209	Chemicals	4043	799.8
210	Iron and steel	5541	1001.5
211	Transportation Equipment	7201	1209.6

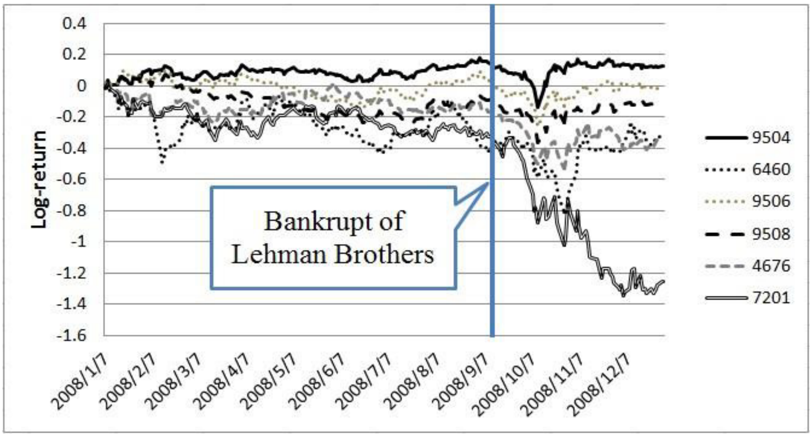

Fig.6. The Top5 stocks of the highest randomness are safer than the stock of the lowest randomness (Code 7201)

(2)High return with high risk

The ranking of randomness in 2008 is shown in Table 4, where the stock of the highest randomness is 8308 that belongs to the sector of Banks, which is not a stable stock. In other words, the empirical rule "high randomness means low risk" does not seem to work. The reason for this may be viewed as the effect of the Lehman shock occurred in September 2008, which made the rest of this year the period of an abnormal fluctuation of stock price. By amputating the abnormal part and using the data just before the end of August in 2008, it is expected to have the normal condition. The result shown in Table 5 indeed shows the normal features, where the stable stock 9506 in Electric/Gas sector is extracted as the stock of the highest randomness, while a relatively unstable stock 7201 in Transportation and Equipment sector shows the lowest randomness. Because 7201 fell down enormously due to the effect of Lehman shock, it is very likely to rise up in 2009. As the result, as shown in Fig 7, 7201 began to rise in March 2009. This is a good example of a high profit accompanied by a high risk. On the other hand, almost of the top five highest randomness stocks perform better than the lowest randomness stock 7201 but 6728. It means that there will be some noise by using the RMT-test to judge the relationship between the randomness and log-return of stock, however, the empirical rule, "high randomness means low risk" is justified.

Table 4. The ranking of randomness by using the tick data of 2008

Rank	Sector	Code	|Error|
1	Banks	8308	28.4
2	Machinery	7004	30.9
3	Transportation Equipment	7211	31.8
4	Electric/ Gas	9502	32.3
5	Electric/ Gas	9508	36.3
	…		
236	Electric Appliances	4902	1604.0
237	Electric Appliances	6954	1611.2
238	Electric Appliances	7752	1646.9

239	Securities	8604	2059.9
240	Shipping	9104	2097.5

Table 5. The ranking of randomness by using the tick data Jan-Aug. ’08

Rank	Sector	Code	|Error|
1	Electric/ Gas	9506	11.6
2	Electric Appliances	6728	13.1
3	Foods	2267	19.1
4	Electric/ Gas	9502	19.1
5	Retail Trade	2685	22.2
		…	
230	Electric Appliances	6762	183.0
231	Electric Appliances	6503	183.7
232	Banks	8306	190.0
233	Securities	8604	222.8
234	Transportation Equipment	7201	257.9

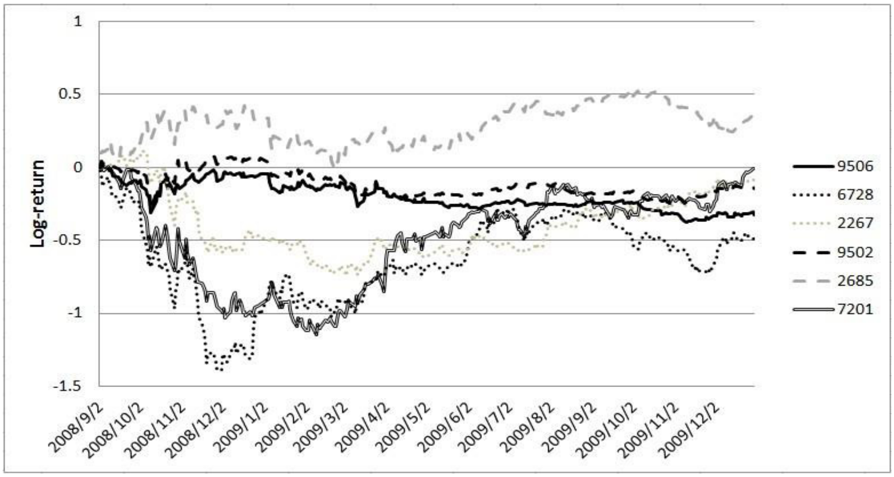

Fig.7. The Top5 stocks of the highest randomness (except 6728) are safer than the stock of the lowest randomness (Code 7201)

According to the result of 2007 and the data before the Lehman shock in 2008, it is shown that the stock which has the highest randomness perform better than the lowest randomness stock, meanwhile, the highest randomness stock is less likely to be affected when there is a financial crisis. The result of 2009 is shown in Table 6 and Fig 8, it is also extract the stock 9509 from Electric /Gas sector as the highest randomness stock, and the almost of the top five highest randomness stocks perform better than the lowest randomness stock 8058 from Wholesale Trade sector in the next year. The empirical rule, "high randomness means low risk” also worked in the stock market of 2009.

Table 6. The ranking of randomness by using the tick data of 2009

Rank	Sector	Code	|Error|
1	Electric/ Gas	9508	25.5
2	Electric/ Gas	9509	28.1
3	Electric/ Gas	9506	31.6
4	Electric/ Gas	9502	33.2
5	Retail Trade	2651	36.4
	…		
225	Nonferrous Metals	5713	1039.5
226	Electric Appliances	4902	1072.4
227	Machinery	6301	1090.9
228	Iron and steel	5541	1128.2
229	Wholesale Trade	8058	1249.5

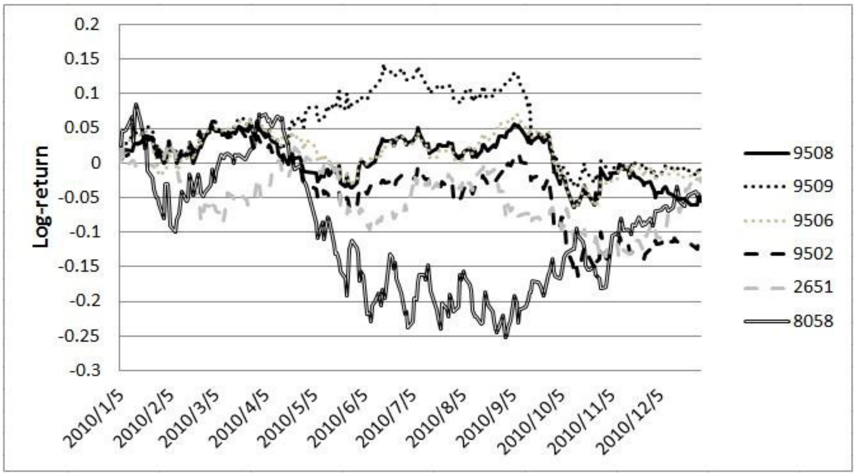

Fig.8. The stock of the highest randomness (Code 9508) is safer than the stock of the lowest randomness (Code 8058)

(3)The randomness of defensive stocks is high

A new classification of related stocks is introduced as shown in Table 7 [7]. The authors focus on the defensive stocks which are Foods, Medicine and Electric/Gas. The defensive stocks are known that they are insensitive to economic activity.

The result in Table 3 shown that the 3 of the top 5 are defensive stocks; Table 5 shown that the 3 of top 5 are defensive stocks; Table 6 shown that 4 of the top 5 are defensive stocks. Therefore, the authors consider that the randomness of defensive stocks is high. On the other hand, all of the 5 stocks which have lowest randomness in Table 6 are Market sensitive stocks and Outer demand stocks, it means that the randomness of Market sensitive is low, that may obtain high return with high risk.

Table 7. Related stock classification

Classification	sectors
Outer demand	「65 : Electrical precision machine」, 「70 : Automobile」
Inner demand	「17 : Construction」, 「88 : Real estate」
Activity sensitive	「30 : Material・Chemistry」, 「60 : Machine」
Defensive	「20 : Foods」, 「45 : Medicine」, 「95 : Electric・Gas」
Consumer	「81 : Retail」, 「90 : Transportation」, 「94 : Telecom/Service」
Interest sensitive	「83 : Bank」, 「85 : Finance」
Market sensitive	「50 : Energy resource」, 「54 : Steel/Metal」, 「80 : Trading/Wholesale」

4. Summary

In this paper, the authors have investigated the relationship between the randomness of the stock price time series and the performance of the stocks in the next period of time, as a new application of the RMT-test. The RMT-test is a new method which can measure the randomness of given long sequences. By using the 3 years data from 2007 to 2009, the authors have discovered that the stocks which have the highest randomness are stable stocks that belong to the sector of the electric/gas power suppliers. This indicates that the suitable stocks to invest under a bear market have higher randomness that belong to the category of 'defensive' stocks, according to the new classification method introduced by Tanaka-Yamawaki, et. al., while the suitable stocks to invest under a bull market have lower randomness that belong to the category of 'market sensitive' stocks in the same classification method[7]. In this paper, we gives out the empirical rule just based on the 3 years data of TOPIX, and we will do further study to prove the rule by using more data in the near future.

References

[1] Yang X., Itoi R., and Tanaka-Yamawaki M.: Testing Randomness by Means of RMT Formula, *Intelligent Decision Technologies, SIST* 2011; **10**: pp.589-596.

[2] Yang, X., Itoi, R. and Tanaka-Yamawaki, M.: Testing randomness by means of Random Matrix Theory, *Progress of Theoretical Physics Supplement* 2012; **194**: pp. 73-83.

[3] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N. and Stanley, H. E.: Universal and Non-Universal Properties of Cross-Correlations in Financial Time Series, *Physical Review Letter* 1999; **83**: pp.1471-1474.

[4] Tanaka-Yamawaki, M., Yang, X. and Itoi, R.: Moment Approach for quantitative evaluation of randomness based on RMT formula, *Intelligent Decision Thechnologies, SIST(Springer)* 2012; **16**: pp. 423-432.

[5] Park, S. and Miller,Random, K.: Number Generators: Good Ones are Hard to Find, *Communication of ACM* 1988; **31**: pp.1192-1201.

[6] Tanaka-Yamawaki, M.: Cross correlation of intra-day stock prices in comparison to Random Matrix Theory, *Intelligent Information Management* 2011; **3**: pp.65-70.

[7] Tanaka-Yamawaki, M.: Extracting Quarterly Trends of Tokyo Stock Market by Means of RMT-PCA, *Advances in Knowledge-Based and Intelligent Information& Engineering Systems*, 2012; DOI: 10.3233/978-1-61499-1-05-2-2028: pp. 2028-2036.