Security levels of stock investment and the Randomness of Price Fluctuation Measured by the RMT-test

Mieko TANAKA-YAMAWAKI, Xin YANG and Yuuta MIKAMORI*

Department of Information and Knowledge Engineering, Graduate School of Engineering,

Tottori University, 101-4, Koyamacho-Minami, Tottori, 68-8662 Japan

*Probizmo., Ltd., Eiko-Ekinan Bldg., 1-3-2 Minamimachi, Izumo, Shimane, 693-0008 Japan

E-mail: mieko@ike.tottori-u.ac.jp

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, 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.

Keyword: Randomness, RMT-test, RMT-PCA, tick-wise stock price

1. The RMT test

The RMT-test[1,2] is a way to measure the randomness of a given time series, of length NL, by comparing the eigenvalue distribution of the N×N cross correlation matrix, whose elements consist of the inner products of a pair taken from N pieces of length L of this time series, to the corresponding theoretical formula derived by the random matrix theory (RMT), called Marcenko-Pastur distribution:

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

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

valid at the limit of N and L going to infinity, keeping Q = L/N as a constant [3].

In order to discriminate subtle difference of randomness, it is more convenient to quantify the randomness as the inverse of the difference between the k-th moment of the eigenvalues

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

with the corresponding theoretical formula obtained from P_{RMT} in Eq. (1).

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

The difference between m_k and μ_k represents the degree of randomness of the data sequence. The authors have chosen to use the 6^{th} 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.

For example, the left hand side of Fig.1 shows the comparison of the above formulas (solid curve) and the eigenvalue distribution (histogram) for the case of pseudo random number generator(prng), in which |Error| = 0.016. The right hand side of Fig.1 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.

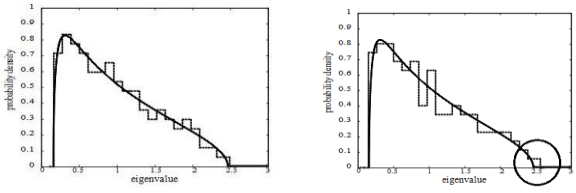

Fig.1 Eigenvalue distribution of the cross correlation matrix(histogram) is compared to the RMT curve (solid curve) in Eq. (1). (left)High randomness given by prng; (right) Low randomness given by the tickwise stock prices of Kao Corporation(code4452).

2. Randomness of Stock Prices

Tick data of stocks in TOPIX 500 from 2007 to 2009 per minute satisfying the appropriate 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. 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 in order to calculate equal time correlation of each stock price. 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

The ranking of randomness in 2007 is shown in Table 2, 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 2 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 2. 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
	...		
210	Iron and steel	5541	1001.5
211	Transportation Equipment	7201	1209.6

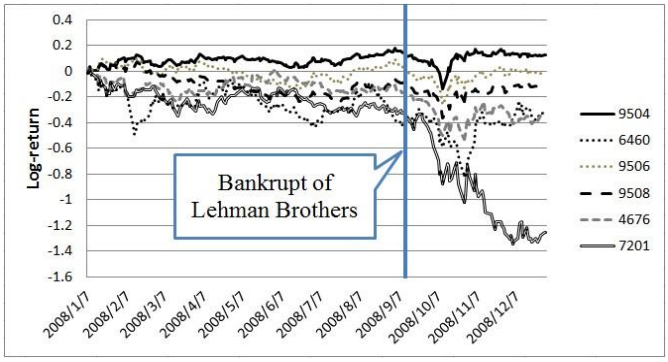

Fig.2. The Top5 stocks of the highest randomness are safer than the stock of the lowest randomness (Code 7201)

However, the ranking of randomness in 2008 does not show the same tendency as the result of 2007. 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 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. The result of 2009 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.

References

[1] 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.

[2] Tanaka-Yamawaki, M., Yang, X., and Itoi, R., "Moment Approach for Quantitative Evaluation of Randomness Based on RMT Formula", Watada, J., et. al.(Eds.) : Intelligent Decision Technologies, SIST Vol. 2, SIST **16**, pp. 423-432, 2012.

[3] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., and Stanley, H.E.,"Random matrix approach to cross correlation in financial data", Physical Review E, Vol. 65, 066126, (2002)