# Xin YANG,<sup>\*)</sup> Ryota ITOI and Mieko TANAKA-YAMAWAKI<sup>\*\*)</sup>

# Department of Information and Electronics, Graduate School of Engineering, Tottori University, Tottori 680-8552, Japan

Random matrix theory (RMT) derives, at the limit of both the dimension N and the length of sequences L going to infinity, that the eigenvalue distribution of the cross correlation matrix with high random nature can be expressed by one function of Q = L/N. Using this fact, we propose a new method of testing randomness of a given sequence. Namely, a sequence passes the test if the eigenvalue distribution of the cross correlation matrix made of the pieces of a given sequence matches the corresponding theoretical curve derived by RMT, and fails otherwise. The comparison is quantified by employing the moments of the eigenvalue distribution to its theoretical counterparts. We have tested its performance on five kinds of test data including the Linear Congruential Generator (LCG), the Mersenne Twister (MT), and three physical random number generators, and confirmed that all the five pass the test. However, the method can distinguish the difference of randomness of the derivatives of random sequences, and the initial part of LCG, which are distinctly less random than the original sequences.

#### §1. Introduction

The random matrix theory<sup>1)</sup> can be used to extract the principal components, by subtracting the random part from the time series with high randomness like stock prices.<sup>2),3)</sup> We propose in this paper a new algorithm of testing the randomness of marginally-random sequences that we encounter in various situations, such as social, economic/financial or medical applications. This method is a straightforward application of the RMT-PCA method<sup>4)</sup> originally developed in order to extract trendy business sectors from a massive database of stock prices. We name this method the 'RMT-test' and examine its effect on several examples of pseudo-random numbers including LCG, and MT, as well as the true random number sequences made by Toshiba, Hitachi, and Tokyo-Electron.

#### §2. Random matrix theory

In applying the random matrix theory, we follow the line of thought that was developed in the course of extracting the principal components of the stock time series in the markets about a decade  $ago.^{2),3}$  Namely, we compare the eigenvalue distribution of the correlation matrix, between N time series of length L, to the corresponding theoretical formula of the eigenvalue distribution<sup>5),6</sup> derived from the random matrix theory in the limit of N and L going to infinity, keeping Q = L/N as a constant.

$$P_{\rm RMT}(\lambda) = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}$$
(2.1)

<sup>\*)</sup> E-mail: yx0709@ike.tottori-u.ac.jp

<sup>\*\*)</sup> E-mail: mieko@ike.tottori-u.ac.jp

## X. Yang, R. Itoi and M. Tanaka-Yamawaki

with

74

$$\lambda_{\pm} = \left(1 \pm \sqrt{\frac{1}{Q}}\right)^2. \tag{2.2}$$

#### §3. Procedure of the RMT-test



Fig. 1. Data structure.

The method of the RMT is outlined as follows.<sup>2),3)</sup> We aim to test the randomness of a long 1-dimensional sequence of numerical data.

• Preparing the data

We prepare a long enough sequence (by using the pseudo-random number generators, or downloading physical random numbers from the web site) and cut it into N pieces of equal length L, then shape them in an  $L \times N$  matrix,  $A_{i,j}$  by placing the first L elements in the first row of the matrix, and the next L ele-

ments in the 2nd row, etc., by discarding the remainder if the length of the sequence is not divisible by L, as shown in Fig. 1. Then we normalize each column of the matrix to have zero mean and single variance,

$$g_{i,j} = \frac{A_{i,j} - \langle A_j \rangle}{\sqrt{\langle A_j^2 \rangle - \langle A_j \rangle^2}}$$
(3.1)

to have the normalized matrix G as follows,

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{L1} & \cdots & g_{LN} \end{pmatrix}.$$
 (3.2)

• Compute the correlation matrix,

$$C = \frac{1}{L} G^T G \tag{3.3}$$

which is symmetric

$$C_{i,j} = C_{j,i} \tag{3.4}$$

by definition and

$$C_{i,i} = 1 \tag{3.5}$$

due to normalization.

- Obtain the eigenvalues of correlation matrix C by numerical calculation.
- Compare the eigenvalue distribution to the corresponding theoretical formula in Eq. (2.1). If the two lines match, that data passes the RMT-test, and if they do not match, it fails the RMT-test. We further quantify the test by adding the next step, in order to discriminate

tiny differences invisible in the visual comparison of the two curves.

• Quantitative evaluation based on the moment method. Compare the *k*-th moment of the obtained eigenvalues

$$m_k = \frac{1}{N} \sum_{i=1}^{N} x_i^k$$
 (3.6)

to the corresponding theoretical formula obtained from  $P_{\rm RMT}$ .

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda.$$
(3.7)

The sample sequence passes the quantitative RMT-test (Quantitative) if the ratio of the moment  $m_k$  over its theoretical value  $\mu_k$  is close to 1. The moments up to 6th can be expressed by the function of Q as follows,

 $\mu_1 = 1, \tag{3.8}$ 

$$\mu_2 = 1 + \frac{1}{Q},\tag{3.9}$$

$$\mu_3 = 1 + \frac{3}{Q} + \frac{1}{Q^2},\tag{3.10}$$

$$\mu_4 = 1 + \frac{6}{Q} + \frac{6}{Q^2} + \frac{1}{Q^3}, \tag{3.11}$$

$$\mu_5 = 1 + \frac{10}{Q} + \frac{20}{Q^2} + \frac{10}{Q^3} + \frac{1}{Q^4}, \qquad (3.12)$$

$$\mu_6 = 1 + \frac{15}{Q} + \frac{50}{Q^2} + \frac{50}{Q^3} + \frac{15}{Q^4} + \frac{1}{Q^5}.$$
 (3.13)

By using those formulas, we evaluate the errors of k-th moments by the deviation of the ratio of the experimental value over the theoretical formula in Eqs. (3.9) to (3.13) from one, as follows,

$$error = m_k/\mu_k - 1.$$
 (3.14)

We can choose the optimal level of error < 5% to judge the randomness of the sequence, based on our experiments.

#### §4. Applications of the RMT-test on random sequences

# 4.1. Determining the reasonable range of N and L

In this section, we determine the reasonable range of N and L. Since the theoretical formula of the eigenvalue distribution  $P_{RMT}$  is derived at the limit of N and L being infinity, we need to choose large enough N and L in order to justify the test. 76

#### X. Yang, R. Itoi and M. Tanaka-Yamawaki



Fig. 2. Errors of the moments  $(k = 1, \dots, 6)$  for  $N = 200, \dots, 500$  for LCG and MT.



Fig. 3. Examples of pseudo-random sequences by LCG passing the RMT-test.

For this purpose, we have applied the test on the data taken from the two popular pseudo random number generators, LCG and MT, for N = 200, 300, 400, 500 at Q = 3, and compared the moments up to the 6th order to the corresponding theoretical formula. We have performed the experiment using 50 different values of seed, of  $SEED = 1, \dots, 50$ , and took the average.

The result is shown in the two figures of Fig. 2, by six lines corresponding the 1st to the 6th moments from the bottom to the top, for LCG (left) and MT (right). In both figures, the errors go down gradually as N increases from 200 to 300, and to 400, then become stable after N reaches the range of 400–500. At N = 500, the errors become smaller than 0.3%. Based on this fact, we justify the value N = 500 being large enough (at least for Q = 3) to apply our RMT-test.

# 4.2. Qualitative evaluation of randomness of pseudo-random generators

The LCG<sup>7</sup> is the most popular pseudo-random number generators, in which the random numbers are generated by the following formula,

$$X_{n+1} = (aX_n + b) \operatorname{mod} M. \tag{4.1}$$

The following parameters are used in the above formula, for rand (),

$$a = 1103515245, \quad b = 12345, \quad M = 2147483648.$$
 (4.2)

We generate a random sequence of length  $500 \times 1,500$ , then cut it into 500 (= N) pieces of length 1,500 (= L) each to make LCG (Q = 3) data. Although LCG is



Fig. 4. Examples of pseudo-random sequences by MT passing the RMT-test.

known to have many problems, the RMT-test cannot detect its off-randomness. As is shown in the left figure of Fig. 3, this data passes the RMT-test safely for Q = 3 (left) with a wide variety of seeds. The right figure of Fig. 3 is a corresponding result for Q = 6 at N = 500, L = 3,000.

The Mersenne Twister  $(MT)^{8}$  is a recently proposed, highly reputed random number generator. The most valuable feature of MT is its extremely long period,  $2^{19937} - 1$ . We test the randomness of MT in the same procedure as above. The result is shown in Fig. 4 for Q = 3, and Q = 6. The MT also passes the RMT-test in a wide range of N and L.

So far, we have seen the two popular pseudo-random sequences pass the RMTtest. We need at this moment some sequences of lower randomness in order to discriminate the level of randomness to be detected by using the RMT-test. The first example is a set of the initial part of random numbers of LCG and MT.

# 4.3. Qualitative evaluation of randomness of physical random number

We test the randomness of some physical random numbers.<sup>9)</sup> Because the physical random numbers have neither regularity nor reproducibility, the forecast of the sequence is impossible (for example: we cannot predict the points when we throw a dice). We use physical random numbers generated by three physical random number generators: Toshiba, Hitachi, and Tokyo-Electron obtained from the homepage of the institute of statistical mathematics, and show the results in Figs. 5, 6 and 7, respectively. The parameters N and L are chosen to be the same as the cases of LCG and MT for the sake of comparison, such that N = 500, L = 1,500 (left) and N = 500, L = 3,000 (right). Note that all the examples pass the RMT-test, for the wide variety of seeds.

# 4.4. Quantitative evaluation of the degree of randomness by means of moments

We have come to the point of discussing the choice of randomness measure. In §4.1, we argued the validity of the RMT-test for the parameter N > 400 for Q = L/N = 3, and 6, based on the comparison of the experimental value over the theoretical value of the k-th moments for k = 1-6, using the last step introduced in Chapter 3. We observed that the errors between the experimental moments and its theoretical counterparts do not converge to zero, but gradually reduce to the 78

X. Yang, R. Itoi and M. Tanaka-Yamawaki



Fig. 5. Examples of physical random numbers by Toshiba passing the RMT-test.



Fig. 6. Examples of physical random numbers by Hitachi passing the RMT-test.



Fig. 7. Examples of physical random numbers by Tokyo-Electron passing the RMT-test.

value less than 0.2-0.3% as N reaches the region of 400-500. Is this because the two pseudo random numbers are not perfectly random, or the RMT formula is not perfectly valid for the finite values of N and L.

In order to answer to this question, we examine the degree of randomness of the physically generated random numbers by the moment method as we introduced in §4.1 and compare the errors to those of the pseudo-random numbers. Since the case of k = 1 is trivial because the average of all the eigenvalues thus equal to one by the normalization condition, Eq. (3.5) and the invariance of the trace under similarity transformations used in the process of diagonalization of the correlation

$\overline{k}$	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	0003(.0009)	.0000(.0010)	0004(.0010)	0004(.0011)	0000(.0008)
3	0007(.0023)	.0001(.0026)	0010(.0027)	0010(.0030)	0002(.0021)
4	0013(.0039)	.0004(.0045)	0017(.0048)	0017(.0053)	0005(.0038)
5	0020(.0058)	.0010(.0067)	0025(.0073)	0022(.0081)	0008(.0057)
6	0026(.0080)	.0018(.0091)	0033(.0101)	0027(.0113)	0010(.0079)

Table I. Q = 3 (L = 1,500) average (standard deviation) of 50 independent tests.

Table II. Q = 3 (L = 1,500) [min.:max.] of 50 independent tests.

$\overline{k}$	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	[0021:.0016]	[0021:.0024]	[0021:.0031]	[0026:.0030]	[0018:.0017]
3	[0051:.0040]	[0052:.0061]	[0058:.0077]	[0064:.0081]	[0047:.0048]
4	[0086:.0065]	[0086:.0109]	[0111 : .0131]	[0107:.0147]	[0080:.0094]
5	[0124:.0104]	[0122:.0164]	[0174:.0191]	[0149:.0226]	[0126:.0149]
6	[0164:.0152]	[0160:.0227]	[0243:.0257]	[0206:.0316]	[0177:.0211]

Table III. Q = 6 (L = 3,000) average (standard deviation) of 50 independent tests.

$\overline{k}$	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	0001(.0005)	0003(.0006)	0001(.0006)	0004(.0010)	0002(.0005)
<b>3</b>	0003(.0014)	0007(.0016)	0004(.0016)	0010(.0021)	0005(.0013)
4	0006(.0026)	0012(.0030)	0008(.0029)	0015(.0035)	0009(.0025)
5	0008(.0041)	0017(.0046)	0012(.0045)	0021(.0051)	0012(.0038)
6	0010(.0059)	0021(.0065)	0016(.0063)	0028(.0069)	0014(.0053)

Table IV. Q = 6 (L = 3,000) [min.:max.] of 50 independent tests.

k	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	[0012:.0009]	[0016:.0010]	[0014:.0011]	[0061:.0011]	[0012:.0008]
3	[0033:.0027]	[0042:.0025]	[0036:.0030]	[0101:.0032]	[0032:.0023]
4	[0060:.0054]	[0075:.0044]	[0064:.0060]	[0150:.0060]	[0059:.0045]
5	[0093:.0088]	[0111:.0072]	[0100:.0098]	[0205:.0092]	[0089:.0078]
6	[0128:.0128]	[0150:.0111]	[0144:.0145]	[0265:.0127]	[0121:.0120]

matrix, we show the results of the k-th moment for  $k = 2, \dots, 6$  in the rest of our discussion. In Table I, we summarize the average values (and the standard deviation in parentheses) of 50 independent experiments for three physical random numbers (Toshiba, Hitachi and Tokyo-Electron) and the two pseudo random numbers (LCG, MT).

The k-th moment ratio for three physical random numbers by the three physical random numbers are compared to the corresponding results of two pseudo-random numbers, LCG and MT (SEED = 1 - 50, N = 500).

Among three physical random generators, Tokyo-Electron generates the most random sequences compared to Hitachi or Toshiba. However, those physical generators are basically less stable compared to the pseudo random generators. In other words, pseudo random generators can produce more uniform sequences with high stability. Moreover, they are deterministic and the entire sequence can be reproduced 80

#### X. Yang, R. Itoi and M. Tanaka-Yamawaki

once the SEED is known together with the algorithm. The results that are shown from Table I to IV tell us that the moment analysis cannot distinguish significant difference among LCG, MT and the three physical random numbers, in accordance with our qualitative test, by which they all pass the RMT-test and the degrees of randomness are indistinguishable. This fact supports the importance of the qualitative test in Figs. 3–7. We conclude that a given sequence passes the RMT-test if it passes the qualitative test, and the corresponding quantitative test by means of moment analysis gives the error less than a few percent for all the moments for k = 1to 6. We next deal with the examples that fail the RMT-test.

# §5. Detecting off-randomness by means of the RMT-test

Table V.	Randomness	of the	e initial	parts	of
pseude	o random nur	nbers.			

k	LCG	MT
2	.0046	0018
3	.0102	0042
4	.0198	0064
5	.0356	0083
6	.0592	0099

In this chapter, we test the offrandomness of two examples by using the RMT test. Our first example is the initial part of LCG sequences and the second example is the log-return sequence frequently used in financial analysis.

#### 5.1. Testing the initial part of LCG

The initial part of LCG is generally believed to have low randomness. In or-

der to quantitatively measure the degree of randomness, we apply the RMT-test on the collection of initial parts of LCG, and compare them with the corresponding data of MT.

We collect the initial 500 numbers generated by iterating the LCG of Eqs. (4.1) and (4.2), starting from various seeds and connect the outputs to serve as out data sequence. As is shown in the left figure of Fig. 8, RMT-test has detected a sign of deviation from RMT formula, for the case of N = 500, and L = 1,500, since some eigenvalues are larger than the theoretical maximum. On the other hand, the corresponding case of the same data without the first 500 numbers after the seeds passes the RMT-test, having all the eigenvalues within the theoretical curve, as shown in the right figure of Fig. 8. For the sake of comparison, we have done the same experiment for another generator, MT, and have confirmed that both the initial and the rest of the sequence pass the RMT-test. The quantitative measure of the off-randomness measured by the first six moments is given in Table V. From this we learn that the errors of moment ratio for the initial 500 elements of LCG are considerably large compared to the corresponding elements of MT.

## 5.2. Testing the randomness of log-return sequences

It is customary to convert the price time series  $p_1, p_2, \dots, p_L$  to the log-return time series  $r_1, r_2, \dots, r_{L-1}$  by means of Eq. (5.1) in the financial analysis, in order to eliminate the unit/size dependence of different stock prices.

$$r_i = \log p_i - \log p_{i-1}. \tag{5.1}$$



Fig. 8. The initial data generated by LCG (MT) fails (passes) the RMT-test.



Fig. 9. RMT-test of the log-return sequences of LCG (left) and MT (right).

Table VI. Errors in RMT-test for the overlapping (left) and non-overlapping (right) log-return sequences.

$\overline{k}$	LCG	MT	7	k	LCG	
2	.1047	.1227		$\overline{2}$	0172	
3	.2578	.3088		3	0436	
4	.4445	.5442	2	4	0744	
5	.6596	.8260	Ę	5	1070	
6	.9092	1.1738	(	6	1400	

However, this process involves the same  $p_i$  for  $r_i$  as well as  $r_{i+1}$ . Because of this, the time series of log-returns loose the randomness that existed in the original price time series and a certain pattern specific to the log-return time series emerges.

In this section, we measure the degree of randomness of such log-return series by using the two pseudo-random generators, LCG and MT, and identify the effect of converting financial time series to the log-return sequence. We compare the results of LCG and MT, by generating the series to make N = 500 and L = 1,500 (Q = 3) and execute the process of steps (1)–(4) in Chapter 3. The results are shown in Fig. 9 and the corresponding moment analysis including the step (5) in Chapter 3 is shown in Table VI (left). We also point out that this effect can be eliminated if we take the non-overlapping log-return by giving up the half of the total elements of  $r_i$ (i = even or odd), in exchange of the length of data L to one half of the original.

#### X. Yang, R. Itoi and M. Tanaka-Yamawaki

-					
$\overline{Q}$	$RMT(4/\sqrt{Q})$	LCG	LCG/RMT	MT	MT/RMT
2	2.82	3.43	1.22	3.43	1.22
3	2.30	2.79	1.21	2.80	1.22
4	2.00	2.40	1.20	2.41	1.21
5	1.78	2.15	1.21	2.15	1.21
6	1.63	1.97	1.21	1.96	1.20
7	1.51	1.81	1.20	1.82	1.21
8	1.41	1.70	1.21	1.70	1.21
9	1.33	1.60	1.20	1.60	1.20
10	1.26	1.50	1.19	1.49	1.18

Table VII. Range of eigenvalues,  $\lambda_+ - \lambda_-$  for LCG and MT are compared to the theoretical value of  $4/\sqrt{Q}$  derived by RMT.

The result of moment analysis is given in Table VI (right). The error compared to the RMT-formula is as large as 100%. This effect results in the expansion of the range of eigenvalues,  $\lambda_+-\lambda_-$  compared to the theoretical range of eigenvalues derived by RMT as  $\lambda_+-\lambda_-=4/\sqrt{Q}$  from Eq. (2.2) and the size of expansion is approximately 20% increase of the theoretical range, as shown in Table VII.

## §6. Conclusion and discussion

#### 6.1. Cumulant analysis

We have so far used the moment analysis in Eqs.  $(3\cdot9)-(3\cdot13)$  for our quantitative test. Often a set of cumulants is used in place of moments. We have derived the corresponding expression of cumulants in terms of parameter Q, up to the 6th order as follows, where  $\kappa_i$  denotes the *i*-th cumulant. In this paper we do not use them for our quantitative analysis because the 6th cumulant gives an extremely large error. However, the following result may become useful for constructing a quantitative test by using the cumulants of the low degree, such as up to 3rd or 5th cumulants.

$$\kappa_1 = \mu_1 = 1,\tag{6.1}$$

$$\kappa_2 = \mu_2 - {\mu_1}^2 = \frac{1}{Q},\tag{6.2}$$

$$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 = \frac{1}{Q^2},\tag{6.3}$$

$$\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 = \frac{1}{Q^3} - \frac{1}{Q^2}, \tag{6.4}$$

$$\kappa_{5} = \mu_{5} - 5\mu_{4}\mu_{1} - 10\mu_{3}\mu_{2} + 20\mu_{3}\mu_{1}^{2} + 30\mu_{2}^{2}\mu_{1} - 60\mu_{2}\mu_{1}^{3} + 24\mu_{1}^{5}$$
  
=  $\frac{1}{Q^{4}} - \frac{5}{Q^{3}},$  (6.5)

$$\kappa_{6} = \mu_{6} - 6\mu_{5}\mu_{1} - 15\mu_{4}\mu_{2} + 30\mu_{4}\mu_{1}^{2} - 10\mu_{3}^{2} + 120\mu_{3}\mu_{2}\mu_{1} - 120\mu_{3}\mu_{1}^{3} + 30\mu_{2}^{3} - 270\mu_{2}^{2}\mu_{1}^{2} + 360\mu_{2}\mu_{1}^{4} - 120\mu_{1}^{6} = \frac{1}{Q^{5}} - \frac{16}{Q^{4}} + \frac{5}{Q^{3}}.$$
 (6.6)

#### 6.2. Discussion

Compared to other conventional methods of testing randomness, the RMT-test that we proposed in this paper can be applied on wide range of numerical data, independent of its data format or types. Moreover, the result is visually presented in a graph that can be grasped intuitively. It is particularly suitable to test the randomness of very long, massive data sequences. No null hypothesis, or other complicated process is required. On the other hand, the method uses a very long data sequence. In order for the RMT-formula to work, we need N strings of length L larger than N, where N is larger than 400. For example, to make N = 500 for Q = 3, we need a data string of length 750,000. Thus the application is limited to the world in which plenty of numerical data can be accumulated.

#### 6.3. Conclusion

In this paper, we proposed a new method of testing randomness, RMT-test, as a by-product of the RMT-PCA that we used to extract trends of stock markets. In order to examine its effectiveness, we tested it on two random number generators, LCG and MT, and three physical random numbers made by Toshiba, Hitachi and Tokyo-Electron. The result shows that all of them pass the RMT-test for a wide range of parameters. Although the physical random numbers by Tokyo-Electron are relatively better than the two other physical random numbers, the degrees of randomness of the three are indistinguishable both in our qualitative test and in our quantitative test. We further tested the validity of our RMT-test on the sequences of low randomness and showed that the RMT-test can detect off-randomness successfully.

#### Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-11-04 on "Econophysics 2011" were useful to complete this work.

#### References

- 1) M. Mehta, Random Matrices, 3rd edition (Academic Press, 2004).
- 2) V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral and H. Stanley, Phys. Rev. E 65 (2002), 066126.
- 3) L. Laloux, P. Cizeaux, J. Bouchaud and M. Potters, Phys. Rev. Lett. 83 (1998), 1467.
- 4) M. Tanaka-Yamawaki, Intelligent Information Management 3 (2011), 65.
- 5) V. Marcenko and L. Pastur, Mathematics of the USSR-Sbornik 1 (1994), 457.
- 6) A. Sengupta and P. Mitra, Phys. Rev. E 60 (1999), 3389.
- 7) S. Park and K. Miller, Communication of ACM 31 (1988), 1192.
- 8) M. Matsumoto and T. Nishimura, ACM Trans. on Modeling and Computer Simulation 8 (1998), 3.
- 9) Y. Tamura, Random Number Library, http://random.ism.ac.jp/random.