Randomness of Tickwise Prices May Forecast the Stock Performance
-A Study on the Randomness of Stock Prices by using the RMT-test-

Xin Yang, Yuuta Mikamori, and Mieko Tanaka-Yamawaki

Department of Information and Knowledge Engineering, Graduate School of Engineering, Tottori University, JAPAN
CONTENT

• Motivation
• About the RMT-test
• Relationship between randomness and safety of stock prices
• Conclusion
Motivation

- **Purpose** ⇒ measuring the security levels of stock investments.

- **Method** ⇒ the RMT-test (quantitative version) as a tool to measure the randomness.
CONTENT

• Motivation
• About the RMT-test
• Relationship between randomness and safety of stock prices
• Conclusion
Rapid progress of information technology has made data collection much easier than ever.

Statistical analysis is performed in various fields. In particular,

**Principal Component Analysis (PCA) by Random Matrix Theory (RMT)**

Use the RMT-PCA to extract useful information
Subtract noisy stocks by the RMT

Extract sectors of strong correlation

Major sectors aggregate (condensed)

Indicator for investment
RMT-PCA: Background of the RMT-test

Yamamoto and Tanaka-Yamawaki (2012)

\[ Q = \frac{L}{N} > 1 \]

\( N \): # of Stocks
\( L \): Data length

Q = 1.31
N = 490
L = 641

Random components
Principal components
The RMT-test

Random Matrix Theory

Measure the randomness of given data sequences
RMT-PCA: Background of the RMT-test

Cross Correlation Matrix

Eigenvalues

Histogram of eigenvalues

Draw $P_{\text{RMT}}$

Compare
Theoretical formula in $\mathcal{P}_{RMT}$

$$L \to \infty, \; N \to \infty ; \; Q = \frac{L}{N}$$

$$\mathcal{P}_{RMT} (\lambda) = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}$$

$$\lambda_\pm = (1 \pm 1/\sqrt{Q})^2$$
Theoretical formula in $P_{\text{RMT}}$
Eigenvalues of real data

Real Data

Correlation Matrix

Eigenvalue Distribution
The RMT test

A tool to measure the randomness of a given data string of any kind
Eigenvalues of real data

$Data_1, \ldots, Data_{(L)}, Data_{(L+1)}, \ldots, Data_{(2L)}, \ldots, \ldots, Data_{(NL)}, \ldots Data$
Eigenvalues of real data

Data_1, \ldots, Data_{(L)} \quad Data_{(L+1)}, \ldots, Data_{(2L)}, \ldots, \ldots, Data_{(NL)} \ldots Data
Eigenvalues of real data

$\text{Data}_1, \ldots, \text{Data}_{(L)}$ $\text{Data}_{(L+1)}, \ldots, \text{Data}_{(2L)}$ $\ldots, \ldots, \text{Data}_{(NL)}$ ...Data

$N$: Num. of data

$L$: Data length

Discard
Eigenvalues of real data

\[ g_{i,j} = \frac{A_{i,j} - \langle A_j \rangle}{\sqrt{\langle A_j^2 \rangle - \langle A_j \rangle^2}} \]

Normalization: \((\text{AV.} = 0, \text{S. D.} = 1)\)
Eigenvalues of real data

\[
C = \frac{1}{L} \begin{pmatrix}
  g_{1,1} & g_{1,2} & \cdots & g_{1,L} \\
g_{2,1} & g_{2,2} & \cdots & g_{2,L} \\
  \vdots & \vdots & \ddots & \vdots \\
g_{N,1} & g_{N,2} & \cdots & g_{N,L}
\end{pmatrix}
\begin{pmatrix}
  g_{1,1} \\
g_{1,2} \\
  \vdots \\
g_{1,L}
\end{pmatrix}
\begin{pmatrix}
  g_{2,1} \\
g_{2,2} \\
  \vdots \\
g_{2,L}
\end{pmatrix}
\cdots
\begin{pmatrix}
  g_{N,1} \\
g_{N,2} \\
  \vdots \\
g_{N,L}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
  1 & C_{1,2} & \cdots & C_{1,N} \\
C_{2,1} & 1 & \cdots & C_{2,N} \\
  \vdots & \vdots & \ddots & \vdots \\
C_{N,1} & C_{N,2} & \cdots & 1
\end{pmatrix}
\]

\[
C_{i,j} = \frac{1}{L} g_i g_j
\]
Eigenvalues of real data

Eigenvalue Distribution
Qualitative evaluation of the RMT-test

Passes the RMT-test

Fails the RMT-test

The same Q
Quantitative evaluation of RMT-test

We employ the moment method in order to compare subtle differences of randomness

\[ \mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{\text{RMT}}(\lambda) d\lambda \]

\[ m_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k \]
Quantitative evaluation of RMT-test

\[ \mu_1 = 1 \]
\[ \mu_2 = 1 + 1/Q \]
\[ \mu_3 = 1 + 3/Q + 1/Q^2 \]
\[ \mu_4 = 1 + 6/Q + 6/Q^2 + 1/Q^3 \]
\[ \mu_5 = 1 + 10/Q + 20/Q^2 + 10/Q^3 + 1/Q^4 \]
\[ \mu_6 = 1 + 15/Q + 50/Q^2 + 50/Q^3 + 15/Q^4 + 1/Q^5 \]
Quantitative evaluation of RMT-test

\[ \mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k p_{\text{RMT}}(\lambda) \, d\lambda \]

\[ m_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k \]

Error \(=|m_k/\mu_k - 1| \times 100\%\)

Smaller error = Higher randomness
CONTENT

• Motivation
• About the RMT-test
• Relationship between randomness and safety of stock prices
• Conclusion
Application to Stock Prices

- Good to compare the randomness of physical random numbers, and pseudo random numbers\(^\text{[1]}\)

- Real application of this method would be the real data of relatively low randomness

Application to Stock Prices

Randomness of tickwise fluctuation

VS.

Profit
Application to Stock Prices

Data:

- **TOPIX500 tick data in 2007-2009 per minute**

- **Tokyo Stock Price Index** commonly known as **TOPIX**, along with the Nikkei 225, is an important stock market index for the Tokyo Stock Exchange (TSE) in Japan

- **TOPIX500 ⇒ Component stocks in the TOPIX Core 30, the TOPIX Large 70 and the TOPIX Mid 400**
## Application to Stock Prices

### Length of tick data (Information/Com.)

<table>
<thead>
<tr>
<th>Code</th>
<th>year</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>9984</td>
<td></td>
<td>66,065</td>
<td>64,608</td>
<td>64,859</td>
</tr>
<tr>
<td>4676</td>
<td></td>
<td>54,350</td>
<td>55,363</td>
<td>58,992</td>
</tr>
<tr>
<td>9684</td>
<td></td>
<td>51,461</td>
<td>53,815</td>
<td>60,018</td>
</tr>
<tr>
<td>4739</td>
<td></td>
<td>48,776</td>
<td>44,136</td>
<td>38,885</td>
</tr>
<tr>
<td>4716</td>
<td></td>
<td>44,165</td>
<td>40,223</td>
<td>34,345</td>
</tr>
<tr>
<td>9401</td>
<td></td>
<td>42,514</td>
<td>47,139</td>
<td>49,568</td>
</tr>
<tr>
<td>4704</td>
<td></td>
<td>40,456</td>
<td>42,495</td>
<td>43,712</td>
</tr>
<tr>
<td>9409</td>
<td></td>
<td>35,096</td>
<td>39,492</td>
<td>42,662</td>
</tr>
<tr>
<td>MAX Length</td>
<td>66338</td>
<td>66338</td>
<td>65945</td>
<td></td>
</tr>
</tbody>
</table>
## Application to Stock Prices

### Data Processing: Part 1

<table>
<thead>
<tr>
<th>time</th>
<th>ticker1</th>
<th>ticker2</th>
<th>ticker3</th>
<th>ticker4</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>502</td>
<td>3810</td>
<td>1902</td>
<td></td>
</tr>
<tr>
<td>09:01</td>
<td>502</td>
<td>3801</td>
<td>1906</td>
<td></td>
</tr>
<tr>
<td>09:02</td>
<td>508</td>
<td>3806</td>
<td>1905</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exclude stocks of short data length</td>
</tr>
<tr>
<td>11:00</td>
<td>521</td>
<td>3788</td>
<td>1910</td>
<td></td>
</tr>
<tr>
<td>12:30</td>
<td>515</td>
<td>3788</td>
<td>963</td>
<td></td>
</tr>
<tr>
<td>12:31</td>
<td>518</td>
<td>3792</td>
<td>953</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:00</td>
<td>522</td>
<td>3820</td>
<td>853</td>
<td></td>
</tr>
</tbody>
</table>
## Application to Stock Prices

### Data Processing: Part 2

| Trading time | AM: 09:00⇒11:00  
| PM: 12:30⇒15:00 |
|---------------|-----------------|
| Data length/day | 272             |
| Data length/year    | 2007 : 245day × 272 - 2*151 = 66338  
|                     | 2008 : 245day × 272 - 2*151 = 66338  
|                     | 2009 : 243day × 272 - 151 = 65945   |
| Data for test     | Stocks of >80%,real prices,  
|                  | (substituted part<20% of total length)  
|                  | 2007 : 211 stocks  
|                  | 2008 : 240 stocks  
|                  | 2009 : 229 stocks  |
Application to Stock Prices

- **Data Processing: Part 3**

- Discard

\[ \text{Data}_1, \ldots, \text{Data}_{(L)} \]

\[ \text{Data}_{(L+1)}, \ldots, \text{Data}_{(2L)} \]

\[ \ldots, \ldots, \text{Data}_{(NL)} \]

\[ \ldots\text{Data} \]
Application to Stock Prices

• Data Processing: Part 3

Randomness of 2009

Data, …, Data, …, Data, Data, …, Data, Data, …, Data, …, Data

relationship

Log-return in 2010
Application to Stock Prices

- Data Processing: Part 3

Randomness of 2009 relationship Log-return in 2010
Application to Stock Prices

- Experiment

Selection of the parameter $Q$
Application to Stock Prices

- **Experiment**
  - $Q>1$ is a condition that the RMT can be applied to data, and a $N \geq 100$ would require at least the lower limit value of $N$
  - The data fixed-length of each year is about 60 thousand
  - Cut the data into pieces by using $N=100$, the $Q$ can be chosen from the range of $[1, 6.6]$, therefore, $Q=2,3,4,5,6$ are the possible integer values for the parameter $Q$
Application to Stock Prices

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.
Application to Stock Prices

Optimal $Q$ is 4
Application to Stock Prices

- Experiment processing

1. \( Q = 4 \)
2. Compute the **randomness** by the RMT-test
3. Sort the stocks according to the randomness to choose the first and the last
4. Investigate the **profit** (log-return) of each stock in the next year
Application to Stock Prices

NIKKEI AVERAGE CHART (2007-2009)

- Year 2007: Significant dropdown in Yen value.
- Year 2008: Steep decline marked by green arrow.
- Year 2009: Recovery shown by red arrow, indicating upward trend.
## Top 10 recording of decline in TOPIX (until 2010)

<table>
<thead>
<tr>
<th>rank</th>
<th>%</th>
<th>TOPIX</th>
<th>y/m/d</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-14.2</td>
<td>1793.90</td>
<td>1987/10/20</td>
<td>Stock prices fall concerned about the Louvre Accord</td>
</tr>
<tr>
<td>2</td>
<td>-9.52</td>
<td>864.52</td>
<td>2008/10/16</td>
<td>Concern for the future of the financial and economic equity by significant setback NY</td>
</tr>
<tr>
<td>3</td>
<td>-8.75</td>
<td>32.32</td>
<td>1953/3/5</td>
<td>Stalin seriously ill</td>
</tr>
<tr>
<td>4</td>
<td>-8.04</td>
<td>899.01</td>
<td>2008/10/8</td>
<td>Fears of the recession of U.S. that Yen appreciation and the NY stock fall</td>
</tr>
<tr>
<td>5</td>
<td>-7.52</td>
<td>806.11</td>
<td>2008/10/24</td>
<td>Significant downward revision of earnings forecast of foreign demand high-tech companies</td>
</tr>
<tr>
<td>6</td>
<td>-7.47</td>
<td>159.33</td>
<td>1970/4/30</td>
<td>Global economic crisis of IOS stock prices</td>
</tr>
<tr>
<td>7</td>
<td>-7.40</td>
<td>746.46</td>
<td>2008/10/27</td>
<td>Appreciation of the yen and downward revision of expected corporate earnings</td>
</tr>
<tr>
<td>8</td>
<td>-7.10</td>
<td>2069.33</td>
<td>1990/4/2</td>
<td>News about Insurance holding bulk sale</td>
</tr>
<tr>
<td>9</td>
<td>-7.10</td>
<td>840.86</td>
<td>2008/10/10</td>
<td>Bankruptcy of life insurance, NY stocks fall sharply</td>
</tr>
<tr>
<td>10</td>
<td>-7.05</td>
<td>889.23</td>
<td>2008/10/22</td>
<td>Concerns about the downturn of high-tech companies</td>
</tr>
</tbody>
</table>

**<Reference> Tokyo Stock Exchange**
Quantitative evaluation of RMT-test

\[ \text{Error} = |\frac{m_6}{\mu_6} - 1| \times 100\% \]

Smaller error = Higher randomness
## Application to Stock Prices

The ranking of randomness by using the tick data of 2009

| Rank | Sectors          | 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           |
Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness

Application to Stock Prices

Highest randomness stock in 2009  ➔  Perform better in 2010

Highest randomness

Lowest randomness
The ranking of randomness by using the tick data of 2007

| Rank | Sectors                          | 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  |
Application to Stock Prices

Highest randomness stock in 2007 → Perform better in 2008

- Highest randomness
- Lowest randomness

Bankrupt of Lehman Brothers

Log-return

2008/1/7 2008/2/7 2008/3/7 2008/4/7 2008/5/7 2008/6/7 2008/7/7 2008/8/7 2008/9/7 2008/10/7 2008/11/7 2008/12/7

9504 7201 NIKKEI
Application to Stock Prices

The stock has the highest randomness is safer
Application to Stock Prices

Highest randomness in 2008
Perform better in 2009?
## Application to Stock Prices

The ranking of randomness by using the tick data of 2008

| Rank | Sectors                     | 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         |
Randomness in all of the 14 industrial sectors

Highest randomness stock in 2008 ➔ Perform better in 2009? ✗

2008 is Abnormal!!!

Lowest randomness

Highest randomness
# Application to Stock Prices

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

| Rank | Sectors               | 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|
Countermeasure

Highest randomness stock in Jan-Aug. ’08 ➞ Perform better in 2009

Highest randomness

Lowest randomness

Countermeasure
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer demand</td>
<td>Stocks sensitive to the OUTSIDE profit</td>
</tr>
<tr>
<td>Inner demand</td>
<td>Stocks sensitive to the INSIDE profit</td>
</tr>
<tr>
<td>Activity sensitive</td>
<td>Stocks sensitive to the economic activity</td>
</tr>
<tr>
<td>Defensive</td>
<td>Stocks insensitive to economic activity</td>
</tr>
<tr>
<td>Consumer</td>
<td>Stocks of consumer’s needs</td>
</tr>
<tr>
<td>Interest sensitive</td>
<td>Stocks sensitive to the interest rate</td>
</tr>
<tr>
<td>Market sensitive</td>
<td>Stocks sensitive to the market trend</td>
</tr>
</tbody>
</table>
## Related stock classification

<table>
<thead>
<tr>
<th>Outer demand</th>
<th>「65 : Electrical precision machine」, 「70 : Automobile」</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner demand</td>
<td>「17 : Construction」, 「88 : Real estate」</td>
</tr>
<tr>
<td>Activity sensitive</td>
<td>「30 : Material・Chemistry」, 「60 : Machine」</td>
</tr>
<tr>
<td>Defensive</td>
<td>「20 : Foods」, 「45 : Medicine」, 「95 : Electric・Gas」</td>
</tr>
<tr>
<td>Consumer</td>
<td>「81 : Retail」, 「90 : Transportation」, 「94 : Telecom/Service」</td>
</tr>
<tr>
<td>Interest sensitive</td>
<td>「83 : Bank」, 「85 : Finance」</td>
</tr>
<tr>
<td>Market sensitive</td>
<td>「50 : Energy resource」, 「54 : Steel/Metal」, 「80 : Trading/Wholesale」</td>
</tr>
</tbody>
</table>
Result: each 3 months

Yamamoto, Tanaka-Y (2012)

Subprime Lending Problem

Lehman Shock

Defensive (D) - Stocks independent of Economic Activity

Consumer (C) - Stocks of consumer’s needs

Interest sensitive (F) - Stocks sensitive to the Interest Rate

2007 2008 2009

1-3 4-6 7-9 10-12 1-3 4-6 7-9 10-12 1-3 4-6 7-9 10-12

0% 20% 40% 60% 80% 100%

1-3 4-6 7-9 10-12 1-3 4-6 7-9 10-12 1-3 4-6 7-9 10-12
Application to Stock Prices

Randomness top 5

- **Stocks independent of Economic Activity**
- **Stocks sensitive to the overseas profit**
- **Stocks sensitive to the domestic profit**
- **Stocks sensitive to the economic activity**
- **Stocks sensitive to the Interest Rate**

<table>
<thead>
<tr>
<th>Year</th>
<th>O</th>
<th>I</th>
<th>A</th>
<th>F</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>100%</td>
<td>80%</td>
<td>60%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>2008</td>
<td>100%</td>
<td>80%</td>
<td>60%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>2009</td>
<td>100%</td>
<td>80%</td>
<td>60%</td>
<td>40%</td>
<td>20%</td>
</tr>
</tbody>
</table>
The stock having the highest randomness is safer

- RMT-test successfully extracted the stock which wasn’t affected by the Lehman shock
- After the ‘big earthquake’ of stock market
  - Used the data before the ‘big earthquake’ only

The sector extracted by RMT-PCA has a high randomness
Thank you