K12is-228
Extracting Market Trends from the Cross Correlation between Stock Time Series

M. Tanaka-Yamawaki, X. Yang, T. Kido, and A. Yamamoto

Abstract In this chapter, the Random Matrix Theory-Principal Component Analysis (RMT-PCA) is applied on daily-close stock prices of American Stocks in NYSE for 16 years from 1994 to 2009 to show the effectiveness and consistency of this method by analyzing the whole data of 16 years at once, as well as analyzing the cut data in various lengths between 2-8 years. The extracted trends are consistent to the actual history of the markets. The authors further analyze the intra-day stock prices of Tokyo Stock Market for 12 quarters extending from 2007 to 2009 and attempted to answer to the two remaining question of the RMT-PCA. The first issue is the number of principal components to examine, and the second issue is the number of eminent elements to examine out of the total N components of the chosen eigenvectors. While the second issue is still open, the authors have found for the first issue that only the second largest principal component is sufficient to examine, based on the comparison of this scenario and the use of the largest ten principal components. This paper argues on this point that the positive elements, and the negative elements, of the eigenvector components individually form collective modes of industrial sectors in the second eigenvector u_2, and those collective modes reveal themselves as trendy sectors of the market in that time period. The authors also discuss on the problem of setting the effective border between the noise and signals considering the artificial correlation created in the process of taking log-returns in analyzing the price time series.

Key words: RMT-PCA, Quarterly Trends, Eigenvector Components, Correlation Matrix

Mieko Tanaka-Yamawaki
Department of Information and Knowledge Engineering
Graduate School of Engineering Tottori University,
e-mail: mieko@ike.tottori-u.ac.jp

1 Introduction

Recently, there have been wide interests on the use of the Random Matrix Theory (RMT) in many fields of sciences [1–10]. In particular, the use of asymptotic formula of the eigenvalue spectrum of cross correlation matrix between independent time series of random numbers [11,12], as a reference to the corresponding spectrum derived from a set of different stock price times series in order to extract principal components effectively in a simple way [13–16], has attracted much attention in the community of econo-physics [17, 18]. The main advantage of this method as a principal component analysis is its simplicity. While the standard Principal Component Analysis(PCA) gives out a way to find the largest Principal Component(PC) and subtract this component from the entire data, and apply the same procedure recursively on the remaining data one by one, the Random Matrix Theory - Principal Component Analysis (RMT-PCA) can process all the ”non-random” components at once by subtracting the RMT formula from the eigenvalue spectrum of a cross correlation matrix. Plerau, et al. [13] was one of the first attempts to apply this technique on stock price time series. By using the daily close stock prices of NYSE/S&P500, they successfully extracted eminent stocks out of massive data of price time series.

However, this method suffers from two difficulties. One is the restriction on the data structure. The entire set of N×T data are needed for analysis, due to the fact that the basic quantity is the cross correlation matrix whose elements are the equal-time inner products between a pair of stocks.

Another difficulty is the restriction of the parameter size. Since the RMT formula is derived in the limit of N and T being infinity, a special care is needed to keep the ranges of the parameters in which the RMT formula is valid.

By using machine-generated random numbers, such as rand(), etc., the authors have tested the validity of the RMT formula in various range of N and T, and have clarified that $N \geqslant 300$, is the safe range unless T is not too close to N, and the validity decreases for smaller N, and the borderline is around $50 < N < 100$. Since the size of stocks dealt in the major markets exceeds 400, the applicability of RMT formula is justified.

Due to the restriction of the methodology to prepare the length of the time series, T, larger than the dimension of the correlation matrix, N, all the data extending to several years had to be combined into a single correlation matrix in [13–16], in which daily-close prices were used. Thus it was difficult to pin-point a short term trend or to compare trends of different time periods.

By employing intra-day (tick-wise) data containing all the transactions made every day, it has become possible to analyze one-year data to compare the result of different years. The authors carried out the same line of study used in [13, 14] by setting up the algorithm of RMT-PCA to be applied on intra-day equal-time price correlations. Based on this approach, the papers [19,20] have shown that this handy methodology works well to extract the trend change of 4 year interval, from 1994 to 2002.

The authors have applied the same algorithm to a wider set of stock price data including daily-close prices of American stocks in the database of S&P500 for 16 years from 1994 to 2009, by cutting the 16 years into 2, 4 and 8 pieces and check the consistency and effectiveness of the proposed methodology in various data lengths. In this paper, the RMT-PCA is applied to a tickwise price data of Tokyo Stock Market from 2007 to 2009, in order to study quarterly trends of the market and attempt to clarify the remaining two technical problems of this algorithm.

There are still some technical problems remaining in the application of the RMT-PCA. One is the number of principal component to be analyzed. It is well known that the first principal component corresponding to the largest eigenvalue of the cross correlation matrix does not give out much information on the trendy sectors, since this mode is almost parallel to the major index of the market made of large-sized popular stocks thus extremely stable [13]. The next largest mode represented by the second eigenvector, u_2 is the major source of information the trendy sectors of that period can be extract from it. Based on the condition $\lambda_i > \lambda_+$, there are 11 to 20 principal components extracted from each data set. Whether u_2 is sufficient for redthe purpose that is to determine the trendy sectors, or some of the remaining states are to be considered is the focus of question.

Another problem is how many elements are to be picked up in order to identify the trendy sectors from the total N dimensional eigenvector, such as u_2. In the previous work of the authors, a fixed number (say 5 or 10) of the largest elements are chosen from each of the positive and negative elements. This point is examined by comparing the use of the fixed number of elements and the fixed accumulative rate.

This paper is organized as follows. After introduction, the methodology of RMT-PCA is summarized in Section 2. The result of daily-close prices of American stocks in the database of S&P500 for 16 years from 1994 to 2009 is shown in Section 3. The result of tickwise price data of Tokyo Stock Market from 2007 to 2009 is given in Section 4, in order to study quarterly trends of the market and attempt to clarify the remaining two technical problems of this algorithm. Then Section 5 is devoted to discuss remaining problems of this methodology.

2 Eigenvalue Problem of Correlation Matrix for Stock Prices

The methodology of the RMT-PCA is outlined as folows. The first step is to prepare the price time series into an N×(T+1) matrix named S, whose i-th row contains the price time series of length T+1. This matrix S is converted into a matrix of log-return as follows

$$r(t) = \log(S(t+\Delta t)) - \log(S(t)) \tag{1}$$

Each string of time series is normalized by

$$x_i(t) = \frac{r_i(t) - \langle r_i \rangle}{\sigma_i} \ (i = 1, \ldots, N) \tag{2}$$

The correlation $C_{i,j}$ between two stocks, i and j, can be written as the inner product of the two log-profit time series, $x_i(t)$ and $x_j(t)$,

$$C_{i,j} = \frac{1}{T}\sum_{t=1}^{T} x_i(t)x_j(t) \tag{3}$$

Here the suffix i indicates the time series on the i-th member of the total N stocks.

The correlations defined in Eq.(3) makes a symmetric ($C_{i,j}$ = $C_{j,i}$), square matrixwhose diagonal elements are all equal to one ($C_{i,i}$) and off-diagonal elements are in general smaller than one ($|C_{i,j}| \leq 1$). As is well known, a real symmetric matrix C can be diagonalized by a similarity transformation $V^{-1}CV$ by an orthogonal matrix V satisfying $V^t = V^{-1}$, each column of which consists of the eigenvectors of C. Such that

$$C_{i,j} = \lambda_k V_k \ (k = 1,\ldots,N) \tag{4}$$

where the coefficient λ_k is the k-th eigenvalue and is the k-th eigenvector.

According to the random matrix theory (RMT, hereafter), the eigenvalue distribution spectrum of C made of random time series is given by the following formula [8,9]

$$P_{RMT}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \ \mathit{where}\ \lambda_\pm = (1 \pm Q^{-\frac{1}{2}})^2 \tag{5}$$

in the limit of $N \to \infty$, $T \to \infty$, $Q = T/N =$ const where T is the length of the time series and N is the total number of independent time series (i.e. the number of stocks considered). This means that the eigenvalues of correlation matrix C between N normalized time series of length T distribute in the following range.

$$\lambda_- < \lambda < \lambda_+ \tag{6}$$

The criterion of the RMT-PCA propoed in this paper is to identify the principal components if the eigenvalues are larger than the upper bound given by the RMT.

$$\lambda_+ < \lambda \tag{7}$$

However, the authors have proved based on extensive numerical analysis using the pseudo random generators that a process of taking the log-return in Eq.(1) adds extra randomness to the data [21–24]. This percolation always occurs and the maximum front of the continuum spectrum extends to about 20% larger than the upper limit λ_+ of RMT. This fact suggests that the upper limit λ_+ is not appropriate to separate the signal from the noise due to the percolation of the random spectrum over

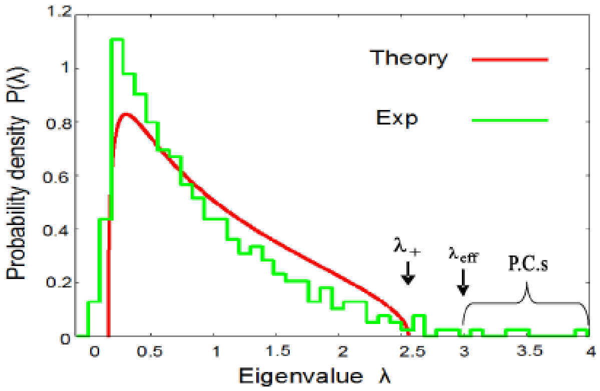


Fig. 1 The states corresponding to the eigenvalues satisfying $\lambda > 1.2\lambda_+ = \lambda_{eff}$ are identified as the principal components by the RMT-PCA.

λ_+ but an effective upper bound λ_{eff}=1.2λ_+. Thus a new criterion is introduced for choosing the principal components

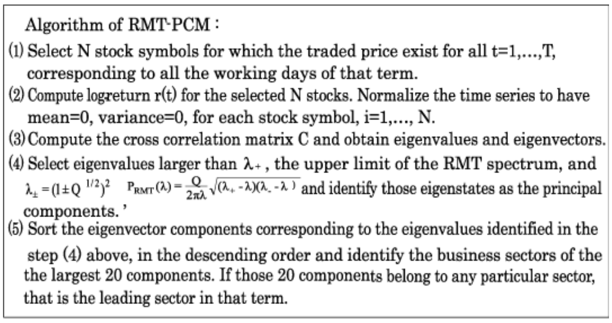
$$1.2\lambda_+ = \lambda_{eff} < \lambda \tag{8}$$

instead of Eq.(7), as illustrated in Fig. 1 above.

3 Trendy Industrial Sectors form the Daily-close Stock Prices

A rectangular matrix of $S_{i,k}$ is constructed by normalizing the N stock returns of the length where i=1, ... ,N represents the stock symbol and k=1, ... ,T represents the traded time of the stocks. The i-th row of this price matrix corresponds to the price time series of the i-th stock symbol, and the k-th column corresponds to the prices of N stocks at the time k. The algorithm to extravt significant principal components is summaried in Fig. 2.

However, a detailed analysis of the eigenvector components has shown that the random components do not necessarily reside below the upper limit of RMT, λ_+, but percolate beyond the RMT due to extra randomness added in the process of computing the log-return in Eq.(1). Based on extensive numerical analysis, this percolation always occurs and the maximum front of the continuum spectrum extends to about 20% larger than the upper limit λ_+ of RMT. This fact suggests that the upper limit λ_+ is not appropriate to separate the signal from the noise due to the percolation of the random spectrum over λ_+ but an effective upper bound λ_{eff} = 1.2 λ_+ about 20% larger than the upper limit λ_+ of RMT. Then λ_+ in the step (4) of the RMT-PCA algorithm in Fig. 2 is to be replaced by λ_{eff}.



```
Algorithm of RMT-PCM :
(1) Select N stock symbols for which the traded price exist for all t=1,...,T,
    corresponding to all the working days of that term.
(2) Compute logreturn r(t) for the selected N stocks. Normalize the time series to have
    mean=0, variance=0, for each stock symbol, i=1,..., N.
(3) Compute the cross correlation matrix C and obtain eigenvalues and eigenvectors.
(4) Select eigenvalues larger than $\lambda_+$ , the upper limit of the RMT spectrum, and
    $\lambda_\pm = (1 \pm Q^{1/2})^2$   $P_{RMT}(\lambda) = \frac{Q}{2\pi\lambda}\sqrt{(\lambda_+ - \lambda)(\lambda_- - \lambda)}$ and identify those eigenstates as the principal
    components. '
(5) Sort the eigenvector components corresponding to the eigenvalues identified in the
    step (4) above, in the descending order and identify the business sectors of the
    the largest 20 components. If those 20 components belong to any particular sector,
    that is the leading sector in that term.
```

Fig. 2 The algorithm to extract the significant principal components in RMT-PCA.

According to the step (4) in the RMT-PCA algorithm in Fig. 2 and, those 14 eigenstates are the principal components, based on. The authors find the business sectors of the companies of 20 largest components in the corresponding eigenvectors. If those components are concentrated in any particular business sector, that sector is defined as the trend makers during that time period. It can be proved mathematically that the eigenvector of the largest eigenvalue is consist of components of the same sign, and the corresponding sectors are not concentrated to a particular sector but distributed to any sectors, because the largest principal component show the global feature of the market thus corresponds to its representative index, such as S&P500, in this case of dealing with American stocks. The eigenvectors of the other eigenvalues have components of both signs. It has been known that the positive components and the negative components belong to the two separate business sectors, if they are strongly concentrated to particular sectors. Summing up those knowledge the authors have, the 2nd principal component reflects the trend of the time period of the data if any concentration of the sectors are observed.

The sectors are classified according to Global Industry Classification Standard(GICS) coding system, that classifies the business sectors of stocks into 10 categories. The authors denote them by a single capital letter, A-J as follows.

A: Energy, B: Materials, C: Industrials, D: Service, E: Consumer Products, F: Health Care, G: Financials, H: Information Technology, I: Telecommunication, and J: Utility.

If taking λ_{eff} instead of λ_+, as it has been explained in the last paragraph of Section 3, then there are 10 eigenstates corresponding to the eigenvalues $\lambda_1 = 74.3, \ldots, \lambda_{10} = 2.41$, actual number of the principal components is less than 14. However, the concentration of business sectors in the eigenvector components occurs only for the 4-5 largest eigenvalues and quickly becomes blur for smaller eigenvalues. Based on this observation, the authors might increase λ_{eff} to the range of λ_{eff} = 2λ, 100% larger than the theoretical criterion. In any case, the difference is irrelevant as long as only several principal components are taken. There are 8 bars corresponding to $v_2(+), v_2(-), v_3(+), v_3(-), v_4(+), v_4(-), v_5(+), v_5(-)$, where $v_k(+)/v_k(-)$ indicates the positive-sign part/negative-sign part of the vector

Table 1 Results for 16, 8, 4 year data (Eigenvalues larger than $2\lambda_+$ are highlighted in bold-Itaric)

	94-09	94-01	02-09	94-97	98-01	02-05	06-09
N	373	373	464	373	419	464	468
T	3961	2015	1946	1010	1002	1006	936
Q	10.6	5.40	4.19	2.71	2.17	2.17	2
λ_+	1.7	2.1	2.2	2.6	2.8	2.8	2.9
λ_1	***74***	***41***	***150***	***37.2***	***53***	***116***	***200***
λ_2	***11***	***13***	***15***	***8.7***	***19***	***14***	***18***
λ_3	***8.8***	***8.8***	***12***	***5.8***	***13***	***13***	***14***
λ_4	***7.7***	***6.9***	***11***	4.6	***9.2***	***9.1***	***8.9***
λ_5	***5.1***	***4.8***	***6.5***	3.3	***6.6***	***6.3***	5.3
λ_6	***4.3***	***4.2***	***5.1***	3.2	***5.8***	5.3	5.0
λ_7	3.3	3.5	3.8	2.8	4.7	4.8	4.4
λ_8	2.9	3.1	3.4	2.6	4.2	4.6	3.5
λ_9	2.5	2.7	3.3	2.4	3.8	4.0	3.2
λ_{10}	2.4	2.2	2.8	2.4	3.8	4.0	3.2
λ_{11}	2.0	2.2	2.4	2.3	2.8	2.9	2.7
λ_{12}	1.9	2.1	2.3	2.3	2.7	2.9	2.5

of k-th principal component, by partitions corresponding to 10 sectors of A-J, and the corresponding eigenvalues and the sign of the components below each bar.

It can be observed from the graphs in Fig. 3 that the sector H (InfoTech) dominates the (+) components of v_2 and the sector J (Utility) dominates the (-) components of v_2.

The result of 8 years data, 1994-2001 and 2002-2009 are shown in Fig. 4, the left figure of which shows the dominance of J (Utility) and H (InfoTech) during the term 1994-2001, and the right figure shows the dominance of A (Energy) and G (Financials) during the term 2002-2009. This means the active sector has changed from J (Utility) and H (InfoTech) to A (Energy) and G (Financials) at the turn of the century.

The results of 4 year data, 1994-1997, 1998-2001, 2002-2005, and 2006-2009 are in Fig. 5, showing the dominance of J (Utility) and H (InfoTech) both in 1994-1997 and 1998-2001, the dominance of A (Energy) and H (InfoTech) in 2002-2005, and A (Energy) and G (Financials) dominance in 2006-2009. The corresponding result of 2 year data is shown in Fig. 6. No clear structure is seen after 2002, except weak dominance of G (Financials) and A (Energy).

The authors have pointed out that the trend of each time period can be successfully depicted by the concentrated business sectors in the positive components and the negative components of the eigenvector corresponding to the 2nd principal components. Although the condition $\lambda > \lambda_+$ dramatically reduces the number of principal components compared to the conventional method of PCA. Moreover, the method proposed in this paper is considerably simple with much shorter in process to extract principal components, which is a great advantage in the case of analyzing the stock market.

The conventional PCA can extract the largest principal component and subtract this element from the entire data, and apply the same procedure recursively on the remaining data one by one. This kind of method requires a lot of computational time and is not suitable for analyzing a system of the large dimension, such as a set of stocks in the market. Another method of PCA uses the eigenvalues of the correlation matrix of times series, which pick up the components whose eigenvalues are larger than one, or the accumulated sum of eigenvalues exceeds 80 percent of the total sum, etc. Neither one is suitable for analyzing the stocks in the market, since the number of principal components thus obtained usually exceeds 100 for N=400-500, while the RMT- PCA has derived the number of principal components in the range of 5-13 in Section 4 in this paper. This point is illustrated in Fig. 9.

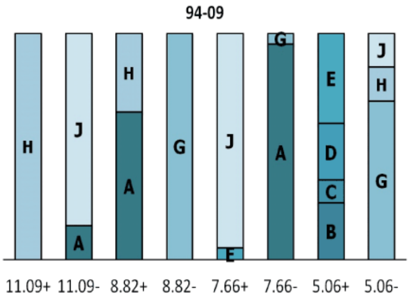


Fig. 3 Trends of 16 years from 1994 to 2009 are shown. The sector H (Information Technology) and J (Utility) are the most eminent sectors in this period.

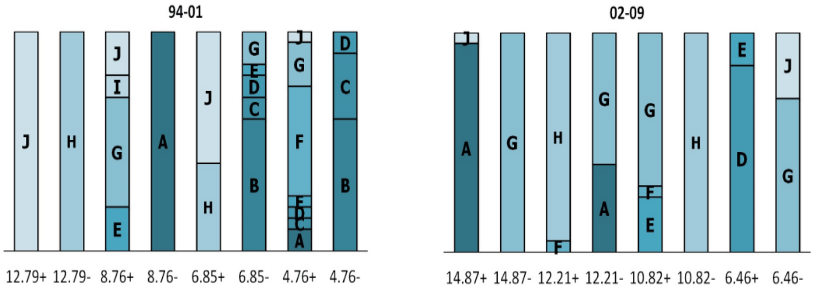


Fig. 4 Trends of 8 years, 1994-2001 (left) and 2002-2009 (right). In 1994-2001, the sector J (Utility) and H (Information Technology) dominance, but in 2002-2009, A (Energy) and G (Financial) dominance the market.

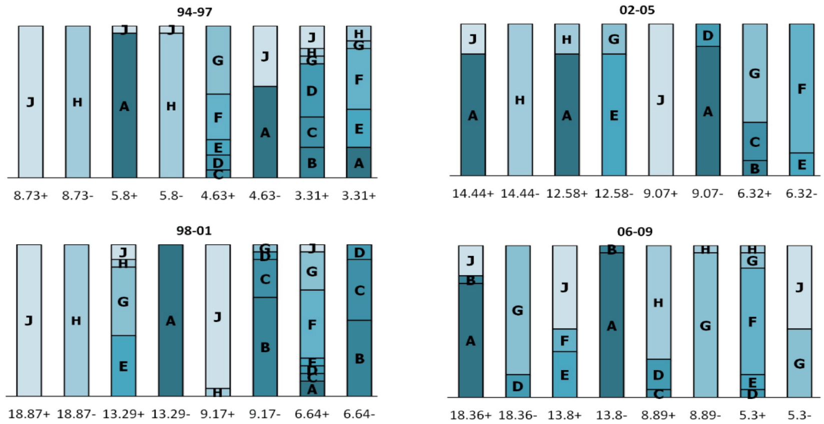


Fig. 5 Trends of 4 years each are shown. Both in 1994-1997 and 1998-2001, J (Utility) and H (Information Technology) dominance, while A (Energy) and H (Information Technology)dominance in 2002-2005 and A (Energy) and G (Financial) dominance in 2006-2009.

4 Trendy Industrial Sectors form the Tickwise Stock Prices

The original tick-wise stock prices are converted to 30 minutes data by selecting the stocks which have at least one trade in the range of each 30 minutes period. For example, the first quarter of the year 2007, from January to March, 2007 had N=486 stocks satisfied this condition and the length of time series of this period was T=642. The numbers of principal components thus computed are listed in the rightmost column of Table 1. Although there are 7-13 principal components whose eigenvalues λ larger than λ_{eff}, for each set of quarterly (or yearly) data, firstly, focus on the second largest eigenvalue λ_2 and its eigenvector u_2, and ignore the rest. Then comparing the above result to the corresponding results of considering all the first ten eigenvectors, in order to show the superiority of the information from u_2, over the noisy results of using other eigenvectors.

First of all, the largest principal component corresponding to the largest eigenvalue λ_1 and its eigenvector u_1 are unfortunately not suitable for extracting trendy sectors. The components of u_1 are almost equally sized around the average value $0.05 = 1/\sqrt{500}$ and do not have any distinguished components, as shown in the first row in Fig. 5. This fact is in common to most markets and is often referred to the ‘market mode’. It is known that this component is strongly correlated to the index consist of dominant and stable stocks such as so called blue-chip stocks.

The authors thus focus on the second largest eigenvalue λ_2 and its eigenvector u_2, which exhibits a certain trend that changes from time to time. Moreover, as shown in the second row of Fig. 5, the positive components aggregate to form a certain collective mode by their internal attractive force, and the negative components do the same, so that both + and - components individually form their own collective modes, which represent temporary trend of active sectors in the market.

The third largest eigenvalue λ_3 and its eigenvector u_3 exhibit the similar feature as the second component in a vague manner, and the fourth eigenvalue λ_4 and its eigenvector u_4 do not show any clear feature and behave more like Gaussian, as shown in the third and the fourth raw of Fig. 5. For the fifth or further eigenstates, the sizes of the N components behave more random and the corresponding histograms reach the Gaussian.

Comparing the first four eigenvectors, it is clear that the second eigenvector u_2 exhibits the existence of two collective modes in the positive and the negative sides in a most clear sense. On the other hand, the components of the first eigenvector u1 are distributed evenly and do not show a sign of aggregation. The components of the fourth or higher eigenvectors are highly random and the distribution is close to Gaussian. The third eigenvector seems transient in between.

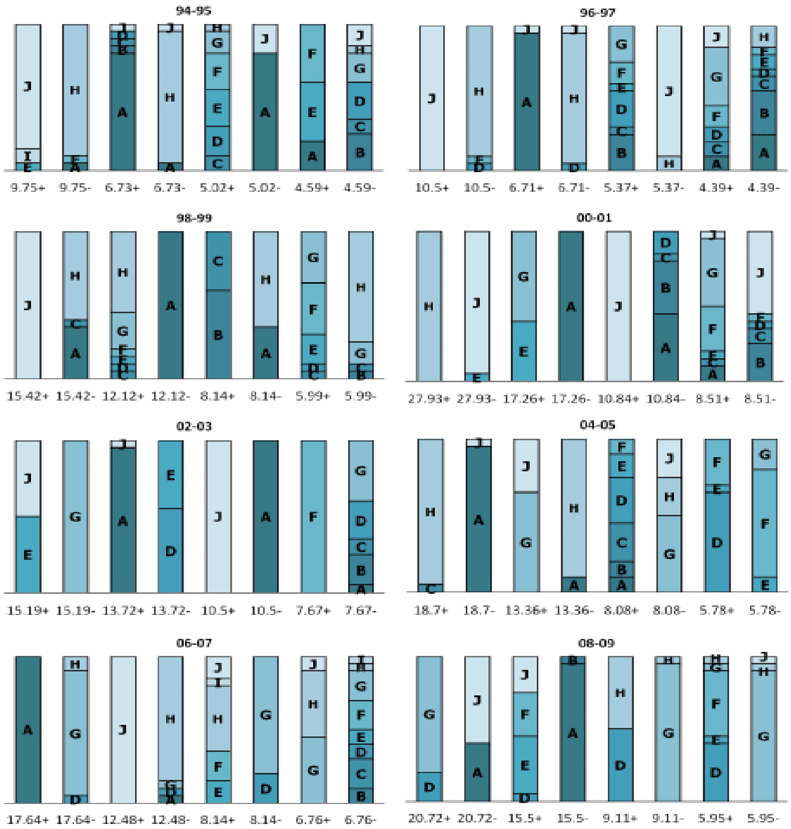


Fig. 6 Trends of 2 years each are shown.The trend change can be observed from J(Utility) and H (Information Technology) dominance towards A (Energy) and G (Financial) dominance.

Table 2 Parameters and the resulting numbers of PCs for the 12 quarters)

YEAR	Quarter	T	N	Q=T/N	λ_+	$\lambda > \lambda_+$	λ_{eff}	$\lambda > \lambda_{eff}$
2007	I	642	486	1.32	3.50	13	4.20	9
	Ⅱ	681	486	1.40	3.40	22	4.08	13
	Ⅲ	681	489	1.39	3.41	18	4.10	12
	Ⅳ	675	492	1.37	3.44	14	4.12	8
2008	I	642	488	1.32	3.50	11	4.20	7
	Ⅱ	681	491	1.39	3.42	14	4.10	9
	Ⅲ	692	492	1.41	3.40	15	4.08	11
	Ⅳ	664	487	1.36	3.45	13	4.14	10
2009	I	642	490	1.31	3.51	11	4.21	7
	Ⅱ	659	486	1.36	3.46	13	4.15	10
	Ⅲ	681	485	1.40	3.40	13	4.08	7
	Ⅳ	670	483	1.39	3.42	15	4.10	10
2007	all	2682	477	5.62	2.02	20	2.43	13
2008	all	2682	480	5.59	2.03	19	2.43	13
2009	all	2655	476	5.58	2.03	16	2.43	10

Based on the above observation, it can be concluded that only u_2 shows a clear sign of the collective modes that make the trendy industrial sectors of each period of time.

5 Trendy Industrial Sectors Extracted from the Collective Modes in the Second Eigenvector u_2

The trendy industrial sectors are identified as the sectors that the distinguishably large elements of the chosen eigenvectors belong to. However, how many of such elements to be taken is not given in any sense. In this paper, the authors followed on this point two different scenarios. One scenario is to take a fixed number of elements from + and - elements each. Fig. 8 shows the result of this scenario for u_2 only. The numbers inside the graphs show the industrial sector according to the codes defined in Table 3. Another scenario is to use the accumulation of large elements in the descending order of the sizes up to 20% of the total amount. Fig. 8 shows the result of this scenario for all the largest ten eigenstates.

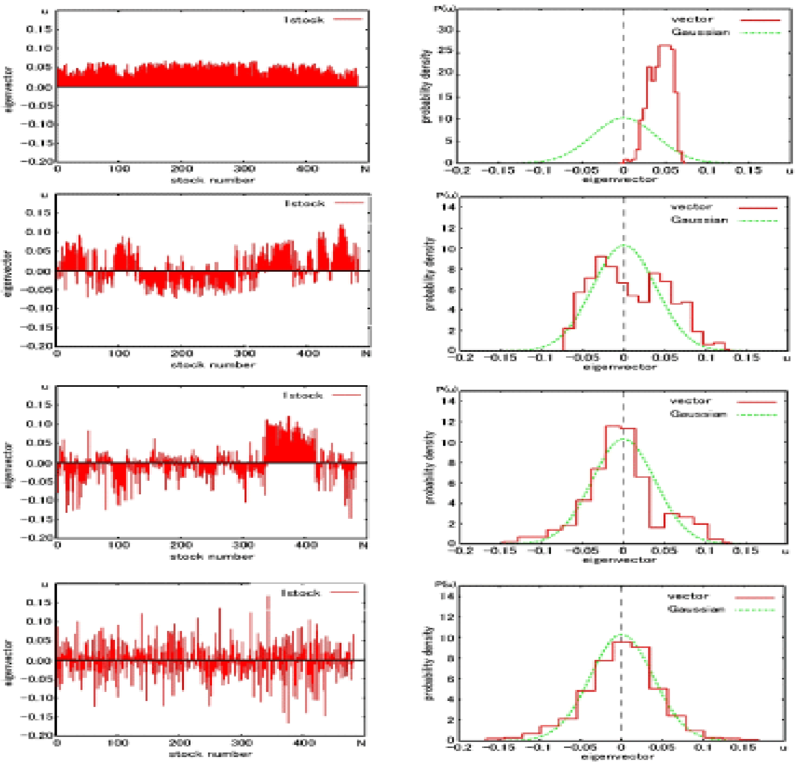


Fig. 7 Eigenvector components (left) and the histograms (right) of u_1 - u_4.

Table 3 The industrial sectors represented by the code numbers

ID:Sector	ID:Sector	ID:Sector
13:Fishery/Agri/Forestry	15:Mineral Mining	17:Construction
20:Food	30:Fiber/Paper	40:Chemistry/Medicine
50:Resources/Material	60:Machine/Elec.Machinery	70:Automobile/Trans.apparatus
80:Commerce	83:Finance/Insurance	88:Real Estate
90:Transport/Telecom	95:Electric/GasPowerSupply	96:Service

6 Conclusion and Discussion

In this paper, it has shown the result of applying the RMT-PCA on 30 minutes traded prices of 4 quarters each in three years from 2007 to 2009 of Tokyo Market, and compared the result on daily data in sixteen years from 1994 to 2009 of S&P.

By analyzing the size distribution of the N components of the first four eigenvectors, the authors found that only the second eigenvector u_2 has a useful feature for the sake of extracting the trendy industrial sectors from their collective modes,

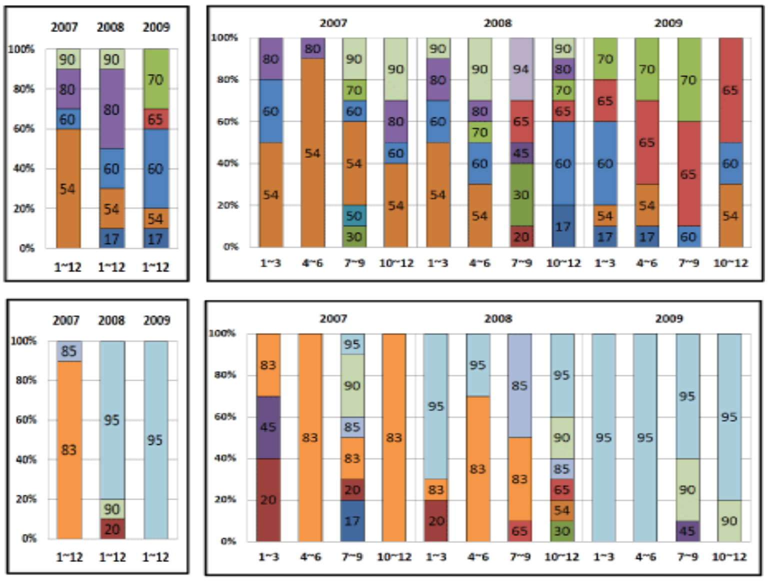


Fig. 8 Trendy sectors in the positive and the negative sectors extracted as collective modes of u_2, obtained from 30 minutes price time series in 2007-2009. The numbers in each bar are the codes of sectors shown in Table 3.

formed independently in the positive parts and the negative parts of the components. This is the conclusion that have reached as to the first of the two unresolved technical problem in the practical application of the RMT-PCA.

As to the second of the unresolved technical problem, the authors have simply compared Fig. 8 where the fixed number of positive and negative elements are selected in the descending order, and Fig. 9 where the accumulation of large elements in the descending order of the sizes up to 20% of the total amount. It is observed that the extracted sectors shown by circles in Fig. 9 coincide with the result of Fig. 8, and no more useful information is offered by Fig. 9 other than noisy details. Thus the authors conclude that the use of u_2 is sufficient to extract trendy sectors, while the number of large elements to consider is inconclusive.

Finally, the authors discuss on the consistency of our result to the actual historical incidence. Both Fig. 8 and Fig. 9 indicate the change of trendy sectors from 83 (banks) in 2007 to 95 (power supply) in 2009. Also a disappearance of major sectors in the third quarter of 2007 and the fourth quarter of 2008 represent the extremely confusing market conditions caused by the sub-prime loan problem in August 2007 and the bankruptcy of Lehman Brothers in October 2008.

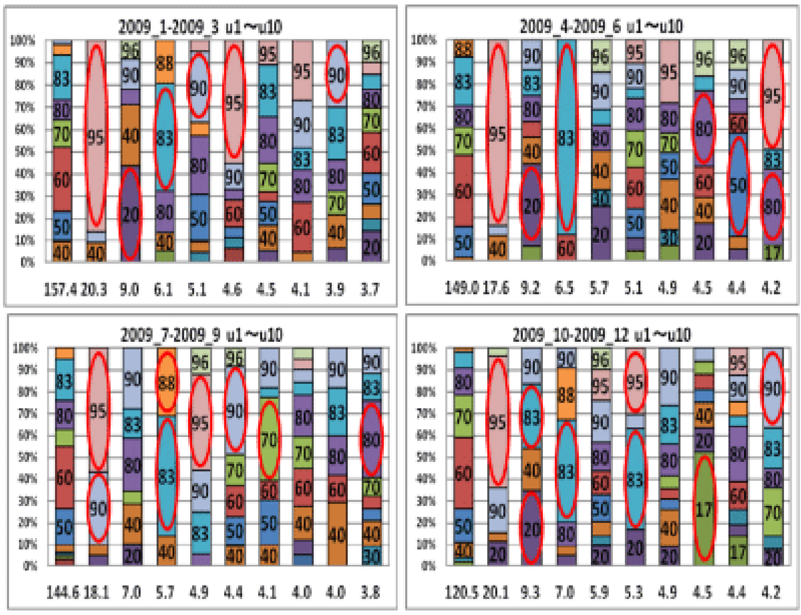


Fig. 9 Noisy result of the top ten eigenstates, u_1-u_{10}, in 2009 are shown for the sake of comparison. The industrial sectors are partitioned in each bar showing 10 eigenstates, ordered from the leftmost bar to the rightmost bar in each figure, with the corresponding eigenvalue below each bar. The first quarter (Jan.-Mar.) to the last quarter (Oct. -Dec.) are shown in four figures from the top to the bottom.

References

1. M. L. Mehta, Random matrices, 3rd edition, Academic Press (2004)
2. A. Edelman and N. R. Rao , Random matrix theory, Acta Numerica, pp.1-65, Cambridge University Press,2005,DOI:10.1017/S0962492904000236
3. Z.D. Bai and J.W. Silverstein , Spectral analysis of large dimensional Random Matrices,Springer (2010)
4. T. Tao and V. Vu and M. Krishnapur, Random matrices: universality of ESD and the circular law, Annals of Probability, Vol. 38, pp. 2023-2065 (2010)
5. C.W.J. Beenakker, Random-matrix theory of quantum transport, Reviews of Modern Physics, Vol. 69, pp. 731-808 (1997)
6. D.A. Kendrick, Stochastic control for economic models, 2nd ed., McGraw-Hill (2002)
7. S.R. Bahcall, Random matrix model for superconductors in a magnetic field, Physical Review Letters, Vol. 77, pp. 5276-5279 (1996)
8. F. Franchini and V.E. Kravtsov, Horizon in Random Matrix Theory, Hawking radiation and flow of cold atoms, Physical Review Letters, Vol.103, 166401 (2009)
9. A. Peyrache, K. Benchenane, M. Khamassi, S.I. Wiener and F.P. Battaglia, Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution, Journal of Computational Neurosience, Vol. 29 ,pp.309-325 (2010)
10. D. Sánchez and M. Büttiker, Magnetic-field asymmetry of nonlinear mesoscopic transport, Physical Review Letters, Vol. 93, 106802 (2004)
11. V.A. Marcenko and L.A. Pastur, Distribution of eigenvalues for some sets of random matrices, Mathematics of the USSR Sb., Vol. 1, No.4 (1967)

12. A.M. Sengupta and P.P. Mitra, Distribution of singular values for some random matrices, Physical Review E, Vol. 60, pp. 3389-3392 (1999)
13. V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley, Random matrix approach to cross correlation in financial data, Physical Review E, Vol. 65, 066126 (2002)
14. V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, Universal and Non-Universal Properties of Cross-Correlations in Financial Time Series, Physical Review Letter, Vol.83, 1471-1474 (1999)
15. L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, Noise dressing of financial correlation matrix, Physical Review Letters, Vol. 83, pp. 1467-1470 (1999)
16. J.-P. Bouchaud and M. Potters, Theory of Financial Risks, Cambridge University Press (2000)
17. R .N. Mantegna and H. E. Stanley, An Introduction to econophysics, Cambridge University Press (2000)
18. H. Iyetomi, Y. Nakayama, H. Aoyama, Y. Fujiwara, Y. Ikeda and W. Souma, Fluctuation-dissipation theory of input-output interindustrial relations, Physical Review E, Vol. 83, 016103 (2011)
19. M. Tanaka-Yamawaki, Extracting principal components from pseudo-random data by using random matrix theory, Lecture Notes in Computer Science, Vol. 6278, pp. 602- 611 (2010)
20. M. Tanaka-Yamawaki, Cross correlation of intra-day stock prices in comparison to random matrix theory, Intelligent Information Management,Vol. 3, pp.65-70 (2011)
21. X. Yang, R. Itoi and M. Tanaka-Yamawaki, Testing randomness by means of RMT formula, Intelligent Decision Technologies, Smart Innovation, Systems and Technologies, Vol.10, pp. 589-596 (2011)
22. X. Yang, R. Itoi and M. Tanaka-Yamawaki, Testing randomness by means of Random Matrix Theory, 2011 Kyoto Workshop on NOLTA pp.1-1 (2011)
23. X. Yang, R. Itoi and M. Tanaka-Yamawaki, Testing randomness by means of Random Matrix Theory, Progress of Theoretical Physics Supplement, No.194, pp. 73-83 (2012)
24. M. Tanaka-Yamawaki, X. Yang and R. Itoi, Moment Approach for quantitative evaluation of randomness based on RMT formula, Intelligent Decision Technologies, Smart Innovation, Systems and Technologies, Vol.16,pp. 423-432 (2012)