

Cross Correlation of Intra-Day Stock Prices in Comparison to Random Matrix Theory

Mieko Tanaka-Yamawaki

Department of Information and Knowledge Engineering, Tottori University, Tottori, Japan

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

Received February 7, 2011; revised March 21, 2011; accepted March 22, 2011

Abstract

We propose and apply a new algorithm of principal component analysis which is suitable for a large sized, highly random time series data, such as a set of stock prices in a stock market. This algorithm utilizes the fact that the major part of the time series is random, and compare the eigenvalue spectrum of cross correlation matrix of a large set of random time series, to the spectrum derived by the random matrix theory (RMT) at the limit of large dimension (the number of independent time series) and long enough length of time series. We test this algorithm on the real tick data of American stocks at different years between 1994 and 2002 and show that the extracted principal components indeed reflects the change of leading stock sectors during this period.

Keywords: Principal Component, Random Matrix Theory, Cross Correlation, Eigenvalues, Stock Market

1. Introduction

Many stock market analysts rely on various technical indicators calculated for individual stocks. However, the difficulty lies in the fact that the optimal time scale varies by time and they can be calculated only after the price time series is recorded. Moreover, the average size of the relevant time scales are not very long, typically shown by the auto-correlation function of the price time series vanish after a few ticks. This fact makes the prediction of future market situation difficult.

Very often, important information to characterize the market can be obtained by observing the change of active business sectors, since multiple stocks in the same business sectors move coherently, independent of the financial status or other parameters of individual stocks.

The method was first proposed and applied on stock prices by Plerou, *et al.* [1,2] on daily close prices (combined with other data) of American stocks and Laloux, *et al.* [3] then applied on daily close prices of Japanese stocks by Aoyama *et al.*

This method uses a result of random matrix theory (RMT) [4-6] on the principal component analysis of the cross correlation matrix of pairs of independent time series. Namely, the method attempts to separate the principal components from the random noises by identifying the random part of the eigenvalue spectrum, to the part

that is identical to the theoretically derived formula [7], out of the eigenvalue distribution of cross-correlation matrix between pairs of stock time series.

Reference [1] argued that the corresponding eigenvector components of the extracted principal components characterizes the eminent business sectors of the years in the period of 1991 to 1998. The reason why the entire 7 years was used for one analysis is the restriction of the methodology requiring $N < T$, where the length T of each time series and the number N of independent time series used for correlation matrix, whose (i, j) element consists of the inner product between the time series of the i -th stock and the j -th stock.

In this paper, we show our challenge of applying the same line of thought as above, on the tick data of the American market during the period of 1994 to 2002. Due to the higher frequency of data points, we can compare the results of different years and describe the historical change of the market in the scenario of this new method of principal component analysis based on RMT spectrum. For the sake of convenience, we name this methodology as RMT-PCA (RMT-oriented PCA).

The rest of this article is constructed as follows. We describe the basic method of RMT-PCA in Chapter 2. Then we show, in Chapter 3, the results of applying the methodology on 1-hour data extracted from the tick data of NYSE and compare the results of computation by

showing the eigenvalue spectrum of 3 years, 1994, 1998 and 2002, compared to the theoretical formula [7] of RMT [6]. We show in this chapter that the truly significant eigenvalues are limited to the first few components, unlike the results of the conventional principal component analysis who counts on the first 80 percent of the accumulated eigenvalues, or other criteria of separating the signal from the noise. In chapter 4, we examine the business sectors which eminent components of eigenvectors of significant principal components belong to. We compare the eminent business sectors of 3 different years, 1994, 1998 and 2002 and show that they indeed reflect the actual list of active stocks during this period of time. Chapter 5 is dedicated to the summary and discussion of this paper.

2. Cross Correlation of Price Time Series

It is of significant importance to extract sets of correlated stock prices from a huge complicated network of hundreds and thousands of stocks in a market. In addition to the correlation between stocks of the same business sectors, there are correlations or anti correlations between different business sectors.

For the sake of comparison between price time series of different magnitudes, we often use the profit instead of the prices [1-5]. The profit is defined as the ratio of the increment ΔS , the difference between the price at t and $t + \Delta t$, divided by the stock price $S(t)$ itself at time t .

$$\frac{S(t + \Delta t) - S(t)}{S(t)} = \frac{\Delta S(t)}{S(t)} \quad (1)$$

This quantity does not depend on the unit, or the size, of the prices which make us possible to deal with many time series of different magnitude. More convenient quantity, however, is the log-profit defined by the difference between log-prices.

$$r(t) = \log(S(t + \Delta t)) - \log(S(t)) \quad (2)$$

Since it can also be written as

$$r(t) = \log\left(\frac{S(t + \Delta t)}{S(t)}\right) \quad (3)$$

and the numerator in the log can be written as $S(t) + \Delta S(t)$,

$$r(t) = \log\left(1 + \frac{\Delta S(t)}{S(t)}\right) \cong \frac{\Delta S(t)}{S(t)} \quad (4)$$

It is essentially the same as the profit $r(t)$ defined on Equation (1). The definition in Equation (2) has an advantage.

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, $r_i(t)$ and $r_j(t)$,

$$C_{i,j} = \frac{1}{T} \sum_{t=1}^T r_i(t) r_j(t) \quad (5)$$

We normalize each time series in order to have the zero averages and the unit variances as follows.

$$x_i(t) = \frac{r_i(t) - \langle r_i \rangle}{\sigma_i} \quad (i = 1, \dots, N) \quad (6)$$

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

The correlations defined in Equation (5) makes a square matrix whose off-diagonal elements are in general smaller than one.

$$|C_{i,j}| \leq 1 \quad (i = 1, \dots, N; j = 1, \dots, N) \quad (7)$$

and its diagonal elements are all equal to one due to normalization.

$$C_{i,i} = 1 \quad (i = 1, \dots, N) \quad (8)$$

Moreover, it is symmetric

$$C_{i,j} = C_{j,i}. \quad (i = 1, \dots, N; j = 1, \dots, N) \quad (9)$$

As is well known, a real symmetric matrix C can be diagonalized by the similarity transformation $V^{-1} C V$ by an orthogonal matrix V satisfying $V^t = V^{-1}$, each column of which consists of the eigenvectors of C .

$$\mathbf{v}_k = \begin{pmatrix} v_{k,1} \\ v_{k,2} \\ \vdots \\ v_{k,N} \end{pmatrix} \quad (10)$$

such that

$$C \mathbf{v}_k = \lambda_k \mathbf{v}_k \quad (k = 1, \dots, N) \quad (11)$$

where the coefficient λ_k is the k -th eigenvalue.

Equation (11) can also be written explicitly by using the components as follows.

$$\sum_{j=1}^N C_{i,j} v_{k,j} = \lambda_k v_{k,i} \quad (12)$$

The eigenvectors in Equation (10) form an orthonormal set. Namely, each eigenvector \mathbf{v}_k is normalized to the unit length

$$\mathbf{v}_k \cdot \mathbf{v}_k = \sum_{n=1}^N (v_{k,n})^2 = 1 \quad (13)$$

and the vectors of different suffices k and l are orthogonal to each other.

$$\mathbf{v}_k \cdot \mathbf{v}_l = \sum_{n=1}^N v_{k,n} v_{l,n} = 0 \quad (14)$$

Equivalently, it can also be written as follows by using Kronecker's delta.

$$\mathbf{v}_k \cdot \mathbf{v}_l = \delta_{k,l} \quad (15)$$

The right hand side of Equation (15) is zero(one) for $k \neq l$ ($k = l$). The numerical solution of the eigenvalue problem of a real symmetric matrix can easily be obtained by repeating Jacobi rotations until all the off-diagonal elements become close enough to zero.

3. RMT-Oriented Principal Component Analysis (RMT-PCA)

The diagonalization process of the correlation matrix \mathbf{C} by repeating the Jacobi rotation is equivalent to convert the set of the normalized set of time series in Equation (6) into the set of eigenvectors.

$$\mathbf{y}(t) = \mathbf{V}\mathbf{x}(t) \quad (16)$$

It can be written explicitly using the components as follows

$$y_i(t) = \sum_{j=1}^N v_{i,j} x_j(t) \quad (17)$$

The eigenvalues can be interpreted as the variance of the new variable discovered by means of rotation toward components having large variances among N independent variables. Namely,

$$\begin{aligned} \sigma^2 &= \frac{1}{T} \sum_{t=1}^T (y_i(t))^2 \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{l=1}^N v_{i,l} x_l(t) \sum_{m=1}^N v_{i,m} x_m(t) \\ &= \sum_{l=1}^N \sum_{m=1}^N v_{i,l} v_{i,m} C_{l,m} \\ &= \lambda_i \end{aligned} \quad (18)$$

since the average $\langle y_i \rangle$ of y_i over t is always zero based on Equation (6) and Equation (17). For the sake of simplicity, we name the eigenvalues in descending order, $\lambda_1 > \lambda_2 > \dots > \lambda_N$.

The theoretical base underlying the principal component analysis is the expectation of distinguished magnitudes of the principal components compared to the other components in the N dimensional space. We illustrate in **Figure 1** the case of 2 dimensional data (x, y) rotated to a new axis $z = ax + by$ and w perpendicular to z , in which z being the principal component and this set of data can be described as 1 dimensional information along this principal axis.

If the magnitude of the largest eigenvalue λ_1 of \mathbf{C} is significantly large compared to the second largest eigenvalue, then the data are scattered mainly along this prin-

cipal axis, corresponding to the direction of the eigenvector \mathbf{v}_1 of the largest eigenvalue. This is the first principal component. Likewise, the second principal component can be identified to the eigenvector \mathbf{v}_2 of the second largest eigenvalue perpendicular to \mathbf{v}_1 . Accordingly, the 3rd and the 4th principal components can be identified as long as the components toward these directions have significant magnitude. The question is how many principal components are to be identified out of N possible axes.

One criterion is to pick up the eigenvalues larger than 1. The reason behind this scenario is the conservation of trace, the sum of the diagonal elements of the matrix, under the similarity transformation. Due to Equation (8), we obtain

$$\sum_{k=1}^N \lambda_k = N \quad (19)$$

which means there exists m such that $\lambda_k > 1$ for $k < k_m$, and $\lambda_k < 1$ for $k > k_m$. This criterion is too loose to use for the case of the stock market having $N > 400$. There are several hundred eigenvalues that are larger than 1, and many of the corresponding eigenvector components are literally random and do not carry useful information.

Another criterion is to rely on the accumulated contribution. It is recommended by some references to regard the top 80 percent of the accumulated contribution are to be regarded as the meaningful principal components. This criterion is too loose for the stock market of $N > 400$, for m easily exceeds a few hundred.

A new criterion proposed in [1-4] and examined recently in many real stock data is to compare the result to the formula derived in the random matrix theory [6].

According to the random matrix theory (RMT, hereafter), the eigenvalue distribution spectrum of \mathbf{C} made of random time series is given by the following formula [7]

$$P_{RMT}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (20)$$

in the limit of

$$N \rightarrow \infty, T \rightarrow \infty, Q = T/N = \text{const.}$$

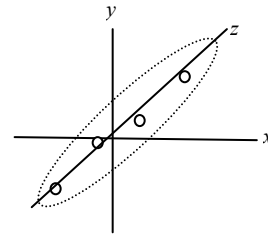


Figure 1. A set of four 2-dimensional data points are characterized as a set of 1-dimensional data along z axis.

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.

$$\lambda_- < \lambda < \lambda_+ \quad (21)$$

where

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \quad (22)$$

are the upper bound and the lower bound of Equation (20). The proposed criterion in our RMT-PCM is to use the components whose eigenvalues, or the variance, are larger than the upper bound given by RMT.

$$\lambda \gg \lambda_+ \quad (23)$$

4. Cross Correlation of Intra-Day Stock Prices

In this chapter we report the result of applying the method of RMT_PCM on intra-day stock prices. The data sets we used are the tick-wise trade data (NYSE-TAQ) for the years of 1994 - 2002. We used price data for each year to be one set. In this paper we mention our result on 1994, 1998 and 2002.

One problem in tickdata is the lack of regularity in the traded times. We have extracted N stocks out of all the tick prices of American stocks each year that have at least one transaction in the 1-hour block including every hour of the days between 10 am to 3 pm. More precisely, among the actual transactions executed between 9:30 to 10:30, the price closest to 10 o'clock is taken as the price at 10am. This provides us a set of price data of N symbols of stocks with length T , for each year. For 1994, 1998 and 2002, the number of stock symbols N as 419, 490, and 569, respectively. The data length T was 1512, six (per day) times 252, the number of working days of the stock market in the above three years. This method can be called as "block-tick" method in comparison to "before-tick" method to be mentioned in Chapter 6. The stock prices thus obtained becomes a rectangular matrix of $S_{i,k}$ where $i=1, \dots, N$ represents the stock symbol and $k=1, \dots, T$ represents the executed time of the stock.

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 .

We summarize the algorithm that we used for extracting significant principal components from 1 hour price matrix in **Table 1**.

Table 1. The algorithm to extract the significant principal components (RMT-PCA) to be applied on tick-wise stock prices.

Algorithm of RMT-PCA:

1) Select N stock symbols for which the traded price exist for all $t = 1, \dots, T$. (6 times a day, at every hour from 10 am to 3 pm, on every working day of the year).

2) Compute log-return $r(t)$ for all the stocks. Normalize the time series to have mean = 0, variance = 1, for each stock symbol, $i = 1, \dots, N$.

3) Compute the cross correlation matrix C and obtain eigenvalues and eigenvectors

4) Select discrete eigenvalues larger than λ_+ in Equation (22), the upper limit of the RMT spectrum, Equation (20), and the continuum spectrum extending the region larger than λ_+ .

Following the procedure described so far, we obtain the distribution of eigenvalues shown in **Figure 2** for the 1-hour stock prices for $N = 419$ and $T = 1512$ in 1994.

The histogram shows the eigenvalues (except the largest $\lambda_1 = 46.3$), $\lambda_2 = 5.3$, $\lambda_3 = 5.1$, $\lambda_4 = 3.9$, $\lambda_5 = 3.5$, $\lambda_6 = 3.4$, $\lambda_7 = 3.1$, $\lambda_8 = 2.9$, $\lambda_9 = 2.8$, $\lambda_{10} = 2.7$, $\lambda_{11} = 2.6$, $\lambda_{12} = 2.6$, $\lambda_{13} = 2.6$, $\lambda_{14} = 2.5$, $\lambda_{15} = 2.4$, $\lambda_{16} = 2.4$, $\lambda_{17} = 2.4$ and the bulk distribution of eigenvalues under the theoretical maximum, $\lambda_+ = 2.3$. These are compared with the RMT curve of Equation (20) for $Q = 1512/419 = 3.6$.

Corresponding result of 1998 data gives the eigenvalue distribution shown in **Figure 3** for $N=490$ and $T = 1512$. In 1998, for $N = 490$, $T = 1512$, there are 24 eigenvalues: $\lambda_1 = 81.12$, $\lambda_2 = 10.4$, $\lambda_3 = 6.9$, $\lambda_4 = 5.7$, $\lambda_5 = 4.8$, $\lambda_6 = 3.9$, $\lambda_7 = 3.5$, $\lambda_8 = 3.5$, $\lambda_9 = 3.4$, $\lambda_{10} = 3.2$, $\lambda_{11} = 3.1$, $\lambda_{12} = 3.1$, $\lambda_{13} = 3.0$, $\lambda_{14} = 2.9$, $\lambda_{15} = 2.9$, $\lambda_{16} = 2.8$, $\lambda_{17} = 2.8$, $\lambda_{18} = 2.8$, $\lambda_{19} = 2.7$, $\lambda_{20} = 2.7$, $\lambda_{21} = 2.6$, $\lambda_{22} = 2.6$, $\lambda_{23} = 2.5$, $\lambda_{24} = 2.5$ and the bulk distribution of eigenvalues under the theoretical maximum, $\lambda_+ = 2.46$. These are compared with the RMT curve of Equation (20) for $Q = 1512/490 = 3.09$.

Similarly, we obtain $N=569$ and $T=1512$ for 2002 data as shown in **Figure 4**, there are 19 eigenvalues, $\lambda_1 = 166.6$, $\lambda_2 = 20.6$, $\lambda_3 = 11.3$, $\lambda_4 = 8.6$, $\lambda_5 = 7.7$, $\lambda_6 = 6.5$, $\lambda_7 = 5.8$, $\lambda_8 = 5.3$, $\lambda_9 = 4.1$, $\lambda_{10} = 4.0$, $\lambda_{11} = 3.8$, $\lambda_{12} = 3.5$, $\lambda_{13} = 3.4$, $\lambda_{14} = 3.3$, $\lambda_{15} = 3.0$, $\lambda_{16} = 3.0$, $\lambda_{17} = 2.9$, $\lambda_{18} = 2.8 = 3.0$, $\lambda_{19} = 2.6$, and the bulk distribution under the theoretical maximum, $\lambda_+ = 2.61$. These are compared with the RMT curve of Equation (12) for $Q = 1512/569 = 2.66$.

However, a detailed analysis of the eigenvector components tells us that the random components do not necessarily reside below the upper limit of RMT, λ_+ , but percolates beyond the RMT limit if the sequence is not perfectly random. Thus it is more reasonable to assume that the border between the signal and the noise is somewhat larger than λ_+ . This interpretation also explains the fact that the eigenvalue spectra always spreads beyond

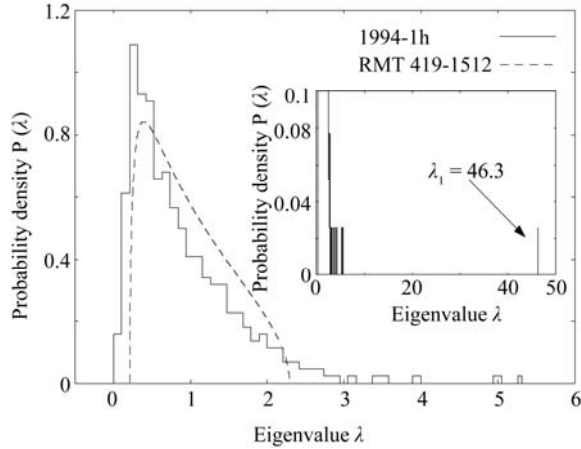


Figure 2. Distribution of eigenvalues of correlation matrix of $N = 419$ stocks for $T = 1512$ data in 1994 compared to the corresponding RMT in Equation (20) for $Q = T/N = 3.6$.

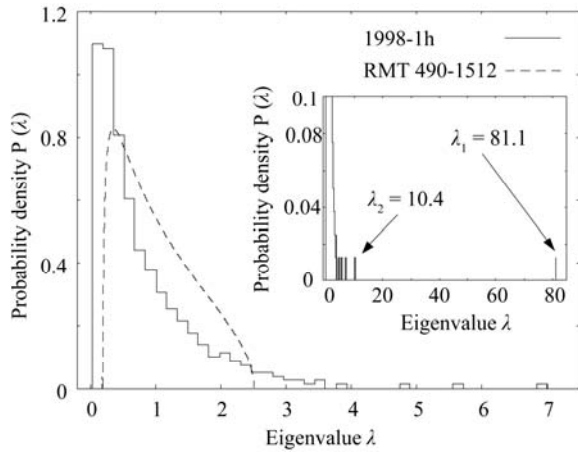


Figure 3. Distribution of eigenvalues of correlation matrix of $N = 490$ stocks for $T = 1512$ data in 1998 compared to the corresponding RMT in Equation (20) for $Q = T/N = 3.09$.

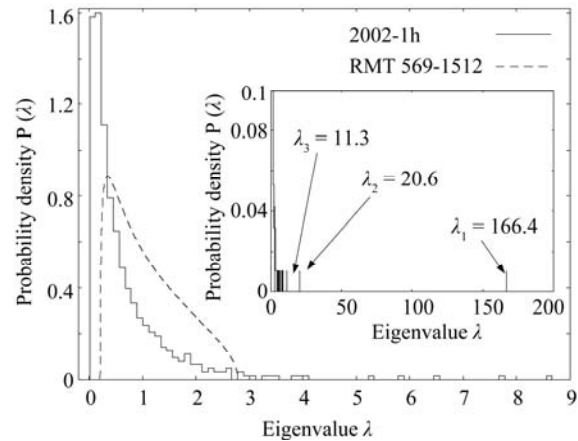


Figure 4. Distribution of eigenvalues of correlation matrix of $N = 569$ stocks for $T = 1512$ data in 2002 compared to the corresponding RMT in Equation (20) for $Q = T/N = 2.66$.

λ_+ . It seems there is no more mathematical reason to decide the border between signal and noise.

However, we have some insight upon this point based on our computer simulations. Namely, the extended continuum over λ_+ occurs even when the time series are random, as long as the return, $r(t)$, instead of the original data, $S(t)$, is used in Equations (1)-(4) in the process of computing the correlation matrix in Equation (5). This causes a characteristic deviation from the randomness and the continuum spectrum extending over λ_+ . We shall discuss on this point in later publication and limit ourselves to make a statement in the step 4 of the RMT-PCA algorithm in **Table 1**.

We return to data analysis in order to obtain further insight for extracting principal components of stock correlation.

5. Eigenvectors as the Principal Components

The eigenvector \mathbf{v}_1 corresponding to the largest eigenvalue is the 1st principal component. For 1-hour data of 1994 where we have $N = 419$ and $T = 1512$ (**Figure 2**), the major components of U_1 are giant companies such as GM, Chrysler, JP Morgan, Merrill Lynch, and DOW Chemical. The 2nd principal component \mathbf{v}_2 consists of mining companies, while the 3rd principal component \mathbf{v}_3 consists of semiconductor manufacturers, including Intel. The 4th principal component \mathbf{v}_4 consists of computer and semiconductor manufacturers, including IBM, and the 5th component \mathbf{v}_5 consists of oil companies. The 6th and later components do not have distinct features compared to the first 5 and can be regarded as random [8].

For 1-hour data of 1998 where we have $N = 490$ and $T = 1512$ (**Figure 3**), the major components of \mathbf{v}_1 are made of banks and financial services. The 2nd principal component \mathbf{v}_2 consists of 10 electric companies, while \mathbf{v}_3 consists of banks and financial services, and U_4 consists of semiconductor manufacturers. The 6th and later components do not have distinct features compared to the first 5 components and regarded as random.

For 1-hour data of 2002 where we have $N = 569$ and $T = 1512$ (**Figure 4**), the major components of \mathbf{v}_1 are strongly dominated by banks and financial services, while \mathbf{v}_2 are strongly dominated by electric power supplying companies, which were not particularly visible in 1994 and 1998.

The above observation summarized in **Table 2** indicates that Appliances/Car and IT dominated the industrial sector in 1994, which have moved toward the dominance of Finance, Food, and Electric Power Supply in 2002.

The advantage in our analysis over that of reference [1] is the use of tick-wise time series. Every year, we have large enough length of $T = 1512$ for all the stocks we

Table 2. Business sectors of top 10 components of 5 principal components in 1994, 1998 and 2002.

ν_k	1994	1998	2002
ν_1	Finance (4), IT (2), Appliances/Car (3)	Finance (8)	Finance (9)
ν_2	Mining (7), Finance (2)	Electric (10)	Food (6)
ν_3	IT (10)	Finance (3)	Electric (10)
ν_4	IT (7), Drug (2)	IT (10)	Food (4), Finance (2), Electric (4)
ν_5	Oil (9)	Mining (6)	Electric (9)

Table 3. Business sectors of top 10 components of 5 principal components for the combined data of 1990-1996 reconstructed from the data given in Table 1 of reference [1].

ν_k	1990 - 1996
ν_1	(Diverse business sectors)
ν_2	Semiconductor (6), Computer (4)
ν_3	Gold (8), Building material (1), Semiconductor/Memory chips (1)
ν_4	Gold (7), Bank (2), Wireless communications (1)
ν_5	(Diverse business sectors)

have used, by taking 6 points per day by selecting the trades nearest to the center of the 6 blocks. This made us possible to analyze each year's data separately and compare different years in the flow of history as in **Table 2**. This was not possible by using only the daily-close price data as in reference [1]. In order to show this fact, we refer the **Table 1** of reference [1] by reconstructing its content of the first five eigenvectors into **Table 3**.

6. Summary and Future Perspectives

In this paper, we propose a new algorithm, RMT-PCA (RMT-oriented PCA) and examined its validity and effectiveness by using the real stock data of 1-hour price time series extracted from the tick-wise stock data of NYSE-TAQ database of 1994, 1998, and 2002. We have shown that this method provides us a handy tool to compute the principal components $\nu_1 - \nu_5$ in a reasonably simple procedure.

We have also tested the method by using two different machine-generated random numbers and have shown that those random numbers work well for a wide range of parameters, N and Q , only if we shuffle to randomize the machine-generated random numbers.

The use of 1-hour price time series made us possible

to compare the results of different years, since there are approximately $T = 1500$ data points per each year for the number of stocks about $N = 500$. Since the Q parameter (T to N ratio) around 3 or larger is the safe area for the usage of Equation (20), we can use 30 minutes, or even 15 minutes data as long as the number of stocks traded at every these time interval are kept large enough to guarantee $Q > 3$. Another possibility is to consider the 'before-tick' data, in which we take the past traded price instead of sticking to the actual traded price during the specified time interval. The advantage of "before-tick" method is twofold. One is the possibility of choosing shorter time interval for one data, such as a quarterly data or a monthly data by yielding larger N and T . Another is the simplicity in dealing with the raw tick-data without taking the time consuming process of extracting actual trades within each block of time interval, as in the "block-tick" method taken in this paper.

7. References

- [1] 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, 2002, pp. 066126. [doi:10.1103/PhysRevE.65.066126](https://doi.org/10.1103/PhysRevE.65.066126)
- [2] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, "Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series," *Physical Review Letters*, Vol. 83, 1999, pp. 1471-1474. [doi:10.1103/PhysRevLett.83.1471](https://doi.org/10.1103/PhysRevLett.83.1471)
- [3] L. Laloux, P. Cizeaux, J. P. Bouchaud and M. Potters, "Noise Dressing of Financial Correlation Matrices," *Physical Review Letters*, Vol. 83, 1999, pp.1467-1470. [doi:10.1103/PhysRevLett.83.1467](https://doi.org/10.1103/PhysRevLett.83.1467)
- [4] J. P. Bouchaud and M. Potters, "Theory of Financial Risks," Cambridge University Press, Cambridge, 2000.
- [5] R. N. Mantegna and H. E. Stanley, "An Introduction to Econophysics: Correlations and Complexity in Finance," Cambridge University Press, Cambridge, 2000.
- [6] M. L. Mehta, "Random Matrices," 3rd Edition, Academic Press, San Diego, 2004.
- [7] A. M. Sengupta and P. P. Mitra, "Distribution of Singular Values for Some Random Matrices," *Physical Review E*, Vol. 60, 1999, pp.3389-3392. [doi:10.1103/PhysRevE.60.3389](https://doi.org/10.1103/PhysRevE.60.3389)
- [8] M. Tanaka-Yamawaki, "Extracting Principal Components from Pseudo-Random Data by Using Random Matrix Theory", KES2010, Cardiff, UK, 2010.
- [9] M. Tanaka-Yamawaki, "Applying Random Matrix Theory to Extract Principal Components of Intra-Day Stock Price Correlations", *Proceedings of the 4th International Conference on New Trends in Information Science and Service Science (NISS2010)*, Vol. 1, 2010, pp. 201-205.