2011 Fourth International Conference on Business Intelligence and Financial Engineering

Analysis of Financial Correlation Matrix Using Random Matrix Theory

Zhu Meiying[1], ZhangBin[2]

1.*Wuhan University, Wuhan, China, 430072*

2.*Nan Yang Institute of Technology, Nanyang,* China, *473004*

zhangbin98168@163.com

***Abstract*—We analyzed the distribution of the eigenvalues of the empirical cross correlation matrix constructed from the data of Chinese stock market and the distribution of the components of some eigenvectors corresponding to certain eigenvalues. Compared with the analytical random matrix theory results, we found that a small part of eigenvalues are out of the range of random matrix theory results and that the distribution of the components of the eigenvector corresponding to the largest eigenvalue is evidently different from that of random matrix theory. These results are similar to the analysis of foreign stock market.**

Keywords: random matrix theory, time series analysis, correlation matrix

I. INTRODUCTION

In recent years, great attention has been drawn to the analysis of some statistical properties of the financial market on the basis of physical concepts and methods. In 1999, US Stanley team used methods of the random matrix theory (RMT) to analyze the time-series correlation matrix of price changes of the largest 1000 US companies on the stock market between 1994 and 1995, and found that the statistical results of most eigenvalues of the eigenvalues spectrum of the correlation matrix agree with the conclusion of the random matrix theory [2,6]. In 2004, Japanese researches used the same method to study the time series of diurnal price changes on the stock market of Tokyo Stock Exchange, and confirmed the universal significance of the results of eigenvalue distribution obtained by the Stanley team [3]. Mainly based on the same method as mentioned above, this paper tries to analyze the behavior of the Chinese stock market. The results show that even though the Chinese market is found to be quite irregular, it is still subject to the universal rules found by Stanley. Our reference data was acquired from the closing prices of 467 typical stocks, which were among the earliest listed A stocks of Shanghai Stock Exchange, between January 2000 and May 2004.

This paper comprises 5 parts. Part 2 below presents the theoretical framework of this paper. Part 3 dwells on data acquisition and data processing. Part 4 contains the analysis results of the abovementioned data of A stocks of Shanghai Stock Exchange and a discussion of these results. Part 5 is a summary of our work.

II. THEORETICAL FRAMEWORK

A. Distribution of eigenvalues of random matrix

According to RMT, the eigenvalue density of random matrix C is [3,5,6]:

$$\rho_c(\lambda)=\frac{1}{N}\frac{dn(\lambda)}{d\lambda} \tag{1}$$

wherein $n(\lambda)$ is the number of eigenvalues of C which are smaller than λ. If C is the $N*T$ order random matrix and $N\to\infty$, $T\to\infty$, $Q=\frac{T}{N}\geq 1$, the specific expression of $\rho_c(\lambda)$ is [2,3,5,6] according to RMT:

$$\rho_c(\lambda)=\frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_{\max}-\lambda)(\lambda-\lambda_{\min})}}{\lambda}$$
$$\lambda_{\min}^{\max}=\sigma^2\left(1+1/Q\pm 2\sqrt{1/Q}\right) \tag{2}$$

wherein $\lambda\in[\lambda_{\min},\lambda_{\max}]$ and σ^2 equals the variance of element M and is normalized to 1. When $Q=1$, the normalized eigenvalue density of M is exactly a representation of the well-known Wigner Semicircle Theory. From Eq.(2) we can see $\lambda_{\max}\rangle\lambda_{\min}\rangle 0$.

B. Distribution of eigenvector components of random matrix

Component i, which corresponds to the eigenvector of eigenvalue λ_α, is expressed as $v_{\alpha,i}$. N is a large number which denotes the dimensionality of the random matrix. u denotes the value of the eigenvector component. $P_N(u)$ denotes the probability density of eigenvector

978-0-7695-4527-1/11 $26.00 © 2011 IEEE
DOI 10.1109/BIFE.2011.15

components between u and $u+du$. Hence we have $\overline{u^2}=1/N$ and [5]

$$P_N(u)=\sqrt{\frac{N}{2\pi}}\exp\left(-\frac{Nu^2}{2}\right) \qquad (3)$$

If we only consider the component distribution of a single eigenvector, then $N=1$ and the above equation is transformed to [2,5,6]

$$P(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right) \qquad (4)$$

C. Construction of financial correlation matrix

The expected revenue of a given investment portfolio P having N assets is defined as:

$$R_p=\sum_{i=1}^{N}p_iR_i$$

$$\sum_{i=1}^{N}p_i=C_p$$

wherein $p_i(i=1,\ldots\ldots N)$ is the amount of capital invested in Asset i, $\{R_i\}(i=1,2,\ldots\ldots,N)$ is the expected revenue of a single asset, R_p is the expected revenue of the investment portfolio, and C_p is the aggregate investment. In order to quantify the correlation, we first calculate the price changes of stocks $i=1,2,\ldots\ldots N$ within a certain period of time:

$$G_i(t)=\ln S_i(t+\Delta t)-\ln S_i(t) \qquad (5)$$

wherein S_i denotes the price of stock i. As different stocks have different levels of lability (standard deviations), we define normalized revenue as:

$$g_i=\frac{G_i(t)-\langle G_i\rangle}{\sigma_i} \qquad (6)$$

wherein $\sigma_i=\sqrt{\langle G_i^2\rangle-\langle G_i\rangle^2}$ is the standard deviation of G_i and $\langle\cdots\cdots\rangle$ denotes the average over a certain period of time. Hence we can calculate the isochronous cross-correlation matrix C. The empirical correlation matrix C may be established with price fluctuation $\delta G_i(t)$ (i denotes Asset i and t denotes time) through the following equation:

$$C_{ij}\equiv\langle g_i(t)g_j(t)\rangle \qquad (7)$$

It is obvious that element C_{ij} denotes the correlation between both securities, its value ranging in the interval of $-1\le C_{ij}\le 1$, wherein $C_{ij}=1$ corresponds to perfect mutual positive correlation and $C_{ij}=0$ corresponds to mutual uncorrelation. Expressed with symbols, Eq.(3) becomes $C=\frac{1}{T}MM^T$, wherein M is an $N*T$ order ROR (rate of return) matrix and M^T is the transposed matrix of M.

III. DATA ACQUISITION AND DATA PROCESSING

The length of the time series from which our data was acquired is T＝1025 ($Q=T/N=1025/467\approx 2.2$, accordingly). Comparison is made between the eigenvalue distribution of the stock correlation matrix constructed with this data and the distribution shown in Eq.(2), and between the eigenvector distribution of the stock correlation matrix and the distribution shown in Eq.(4). The comparison is based on the assumption that: the financial correlation matrix is purely random, i.e. informationless. This assumption is the precondition for valid comparison. If our assumption is justified, then both results should agree with each other by comparison. If they disagree or fail to completely agree, it means that the correlation matrix is not random or not completely random and contains some useful information.

Fig.1 shows the results of comparison between the eigenvalue distribution of the correlation matrix established with this data and the eigenvalue distribution of RMT. Fig.2 shows the results of comparison between the distribution of eigenvector components corresponding to the eigenvalues within the range of RMT and the distribution of eigenvector components of RMT. Fig.3 shows the results of comparison between the component distribution of the maximum eigenvector and the distribution of eigenvector components of RMT.

IV. RESULTS, ANALYSIS, AND DISCUSSION

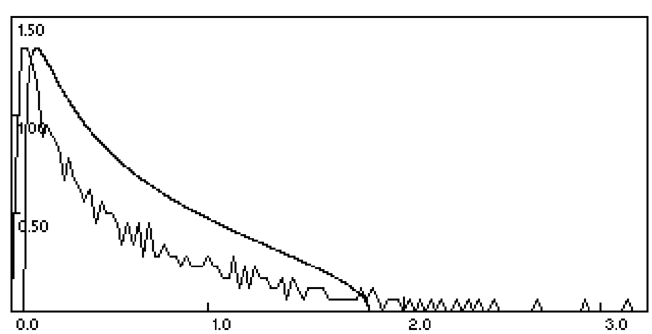

Figure 1. The smooth curve stands for the RMT distribution. The jagged curve stands for the eigenvalue density distribution derived from the time series of 467 A stocks of Shanghai Stock Exchange. The minimum eigenvalue of the random matrix of adopted parameters is 0.0687 and its maximum eigenvalue is 1.8236. The minimum eigenvalue of the financial correlation matrix is 0.0319 and the maximum eigenvalue of the financial correlation matrix is: 162.6429; 89.2934％ of the financial correlation eigenvalues are within the range of RMT.

From Fig.1 we can see that the maximum eigenvalue λ_1 is 68.34 times of the predicted $\lambda_{\max}$, much greater than the maximum eigenvalue of the corresponding RMT. According to the theory of investment, this eigenvalue corresponds to the market. Therefore the above-mentioned "pure random" assumption evidently disagrees with λ_1. A reasonable explanation is that those correlation matrix components that are orthogonal to the market are pure noise. This means that by deducting the contribution of $\lambda_{\max}$ from $\sigma^2 = 1$, we will have $\sigma^2 = 1 - \lambda_{\max} / N = 0.65$. The jagged curve in Fig.1 shows the result corresponding to empirical distribution. Admittedly, some eigenvalues are greater than $\lambda_{\max}$ and may contain some information that reduces the variance of the effective random part of the correlation matrix. The eigenvalue range of RMT covers more than 89% of the eigenvalues of the empirical correlation matrix, and about 11% of these values are indeed outside the theoretical boundaries. And some eigenvalues are greater than $\lambda_{\max}$ or less than $\lambda_{\min}$. For the largest eigenvalues that are greater than $\lambda_{\max}$, we may adjust σ^2 as an adjustable parameter until a better σ^2 is achieved. Those eigenvalues that are less than $\lambda_{\min}$ have little effect on the value of σ^2 after being divided by N.

The assumption we have made above does not lead to the results we expected, indicating that the financial correlation matrix we are considering is not completely random.

The results shown in Fig.1 are similar to the conclusions made for the stock markets of developed countries such as USA [6]. By approximation, we can distinguish information from noise by correcting the theoretical boundaries determined by the density part that covers most of the eigenvalues.

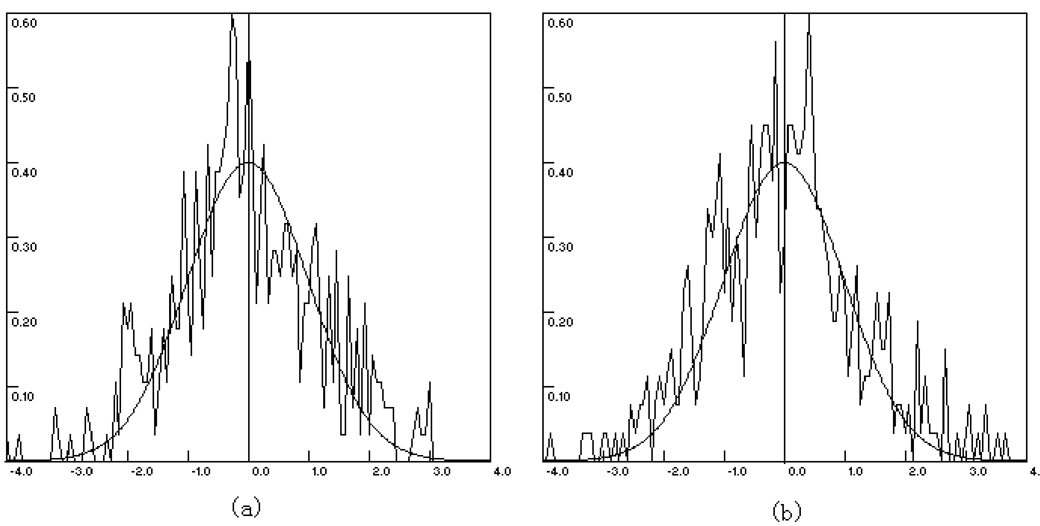

Figure 2. Comparison between Porter-Thomas distribution (smooth curves) and component distributions of any two eigenvectors corresponding to eigenvalues within $[\lambda_{\min}, \lambda_{\max}]$ The theoretical minimum and maximum eigenvalues of the random matrix of selected parameters are 0.0687 and 1.8236, respectively. Fig.2(a) and Fig.2(b) are respectively the component distribution diagrams of the 102nd and the 248th eigenvectors, which respectively correspond to eigenvalues 0.1597 and 0.4643.

Fig.2 shows the component distribution of an eigenvector that corresponds to an eigenvalue within RMT. From the diagram we can see that the distribution of eigenvector components is highly similar to Porter-Thomas distribution (as eigenvectors have been normalized, we amplify each component by 4 times to facilitate comparison with Porter-Thomas distribution. The same method is also used for Fig.3).

From Fig.1 and Fig.2 above we can see that the smaller eigenvalues are actually of a random nature. They are very close to the minimum eigenvalues of RMT and most sensitive to "noise". And their corresponding eigenvectors are exactly those that determine the investment portfolios having less risk. With the RMT method, we can discriminate those eigenvalues and eigenvectors that contain true information from those unstable ones that do not contain any useful information in the correlation matrix. On the other hand, the empirical financial correlation matrix is not completely random. Compared with the completely random matrix, it has some deviations which may contain true and useful information.

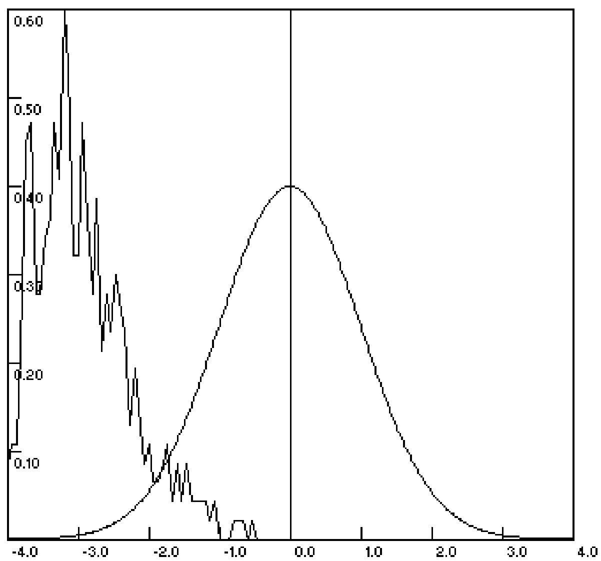

Figure 3. Comparison between Porter-Thomas distribution and component distribution of eigenvector corresponding to maximum eigenvalue The theoretical minimum and maximum eigenvalues of the random matrix of selected parameters are 0.0687 and 1.8236, respectively. The eigenvalue of the financial correlation matrix that corresponds to the diagram is 162.6429.

Figure shows the component distribution of the eigenvector that corresponds to the maximum eigenvalue. It is significantly different from the theoretical distribution of the RM eigenvalue and as a result, the actual result we have is also different from our "informationless" assumption. According to the theory of investment, it should reflect a common factor that influences all assets: the market. This is exactly the underlying principle for the simple single-factor β model in financial applications.

V. CONCLUSION

89% of all the eigenvalues of A stocks of Shanghai Stock Exchange are covered by theoretical equation (2). It is then certain that these eigenvalues are "informationless". So the rest eigenvalues, which takes up a portion of about 11%, should carry some true and useful information. This is a very helpful method for discerning the correlation between different types of financial assets and this method has some potential functions in risk management and the optimization of investment portfolios. While selecting investment portfolios, we can use a random matrix to screen out the random-natured eigenvalues in the financial correlation matrix in order to derive a better investment portfolio. From these results, we can easily see that Markowitz's solution to the optimization of investment portfolios, which is solely based on a correlation matrix derived from history data, is inadequate because its small eigenvalues (which determine which investment portfolios are less risky) are full of noise. This matrix has to be corrected in order to eliminate the effect of noise and thus maximize the profitability of the investment portfolios determined by it. To eliminate the noise, we associate each eigenvector with eigenvalues whose "noise band" is a constant, and thus select those eigenvalues that are always consistent with the traces of the initial correlation matrix.

We have found that our results coincide with the conclusions drawn from the data of the US and Japanese financial markets by using RMT. Specifically, our Fig.1 is basically identical to Fig.1 in Ref.6 and Fig.1 in Ref 3, our Fig.2 is basically identical to Fig.2 in Ref 6, and our Fig.3 is similar to the thumbnail in Fig.2 of Ref.6. This coincidence indicates that RMT is of universal significance for all financial markets.

REFERENCES

[1] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud,and Marc Potters. Noise Dressing of Financial Correlation Matrices. The American Physical Society，1999,83(7):1467~1470

[2] Akihiko Utsugi, Kazusumi Ino, and Masaki Oshikawa1. Random matrix theory analysis of cross correlations in financial markets. The American Physical Society，2004，70(2): 026110-1~026110-11

[3] Vasiliki Plerou, Parameswaran Gopikrishnan,Bernd Rosenow, Luı´s A. Nunes Amaral,and H. Eugene Stanley. Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series. The American Physical Society, 1999,83(7):1471~1474

[4] Li Ping, Wang Binghong, and Quan Hongjun. Several Fundamental Issues and Research Progress of Financial Physics (I). Physics, 2004, 1 (6): 28~33

[5] Li Ping, Wang Binghong, and Quan Hongjun. Several Fundamental Issues and Research Progress of Financial Physics (II). Physics, 2004, 3 (8): 205~212