Comparing TEPIX as an emerging market with efficient market by

Random Matrix Approach

A.Namaki[a] alinamaki1@yahoo.com

{Faculty of Management, Tehran University, PO Box 14819-57711, Tehran , Iran}

G. R. Jafari g_jafari@sbu.ac.ir

{Department of Physics, Shahid Beheshti University, G.C., Evin, Tehran 19839, Iran}

R.Raei rraei@ut.ac.ir

{Faculty of Management, Tehran University, Tehran, Iran}

We analyze cross-correlation between return fluctuations of different stocks by using random matrix theory (RMT). We test the statistics of eigenvalues of cross-correlation (C) between stocks of Tehran price index (TEPIX) as an emerging market in the period of 1 April 2005 to 1 April 2010 and compare these with a mature market (US market). According to the "null hypothesis" ─ a random correlation matrix constructed from mutually uncorrelated times series─ the deviation from the Gaussian orthogonal ensemble of RTM is a good criterion. We find that a majority of the eigenvalues of C fall within the bulk(RMT bounds between λ_- and λ_+) for the eigenvalues of the random correlation matrices. Further, we find that the distribution of eigenvector components for the eigenvectors corresponding to the largest deviating eigenvalues, display systematic deviations from the RMT prediction. We analyze the components of the deviating eigenvectors and find that the largest eigenvalue corresponds to an influence common to all stocks. Our analysis of the remaining deviating eigenvectors shows distinct sectors, whose identities corresponds to the structure of the Iran business environment. Market of Iran is in the shadow of government.

Key words: RMT, eigenvalue , Correlation, emerge market, mature market

[a] Corresponding author

I. INTRODUCTION

Quantifying cross-correlation between different assets is a topic of interest for understanding the economy as a complex system and for practical reasons such an asset allocation and portfolio-risk optimization [1-4]. Here, we analyze cross-correlation between stocks by applying random matrix approach, developed in the context of complex quantum systems where same as stock markets, the precise nature of the interactions between subunits are not known.

In order to do this, we first calculate the price change of stocks i=1,…,.N over a time scale Δt

$$G_i(t) = lnS_i(t) - lnS_i(t - \Delta t)\ , \qquad (1)$$

Where $S_i(t)$ denotes the price of stock i at time t. We define a normalized return in order to standardize the different stock volatilities. We normalize G_i with respect to its standard deviation σ_i As follows[5]:

$$g_i = \frac{G_i(t) - < G_i >}{\sigma_i} \qquad (2)$$

where $\sigma_i = \sqrt{< G_i^2 > - < G_i >^2}$, and <…> denotes a time average over the period studied. We then compute the equal-time cross-correlation matrix C with elements

$$C_{ij} = < g_i(t) g_j(t) > \qquad (3)$$

The elements of C are limited to the domain [-1,1], where C_{ij} =1 defines perfect correlation, C_{ij} =-1 corresponding to perfect anti-correlation and C_{ij} =0 corresponding to uncorrelated pairs of stocks.

This is difficult to analyze the significance and meaning of the empirical cross-correlation coefficients (C_{ij}),because of non-stationary cross-correlation as a result of market conditions and "measurement noise" as a result of finite length of time series available for estimating the C_{ij} .

These problems induce us using a way to identify which stocks remained correlated (on the average) in the time period studied? To answer this question, we test the statistics of C against the null hypothesis of a random correlation matrix constructed from mutually uncorrelated time series. If the properties of C conform to those of a random correlation matrix, then it follows that the contents of the C are random and deviations of the properties of C from those of a random correlation matrix convey information about genuine correlations. Thus, our goal is separating the contents of C into two groups: (a) the part of C that conforms to the properties of RMT ("noise") and (b) the part that deviates ("information") [5,19] . For mature markets have been done many researches in this area, and have found good agreement with this hypothesis [5,19]. In this paper we are trying to calculate the RMT properties for Iran stock market and so, we will be able to comparing this market with mature markets [22] .

This paper is organized as follows: in Section II we review RMT and discuss about it. Section III contains a brief description of the data analyzed. Section IV discusses the statistics of cross-correlation coefficients and comparing this with mature markets. Section V discusses the eigenvalue distribution of C and compares with RMT results. Section VI contains a detailed analysis of the contents of eigenvectors that deviate from RMT. Finally, Section VII contains some conclusion states.

II. Random Matrix Theory

RMT was developed in the context of nuclear physics by Wigner, Dyson, Mehta, and others in order to explain the statistics of energy levels of complex quantum systems [6-10]. The success of random matrices lies in the universality regime of the eigenvalue statistics. There is compelling evidence that when the size of the matrix is very large then the eigenvalue distribution tends, in a certain sense, towards a limiting distribution. This only depends on the symmetry properties of the matrix and is independent of the initial probability law imposed on the matrix entries.

Some researches [11, 12] applying RMT methods to analyze the properties of C show that the most of the eigenvalues of C agree with RMT predictions, suggesting a considerable degree of randomness in the measured cross-correlations. It is also found that there are deviations from RMT predictions for the largest eigenvalues.

We analyze the components of the deviating eigenvectors and find that the largest eigenvalue corresponds to an influence common to all stocks.

By using the inverse participation ratio, we analyze the eigenvectors of C and find large values at both edges of the eigenvalue spectrum — suggesting a “random band” matrix structure for C.

III. DATA ANALYZED

We analyze "Rah Avard Novin" database, that covering all transactions of securities of Tehran Stock Exchange [13]. We extract from this database time series of prices of the 325 stocks of Tehran Stock Exchange, on the starting date April 1, 2005. We analyze daily change of this database over a period of 1291 consecutive trading days in 2005-2010[14]. From this, we form L=1291 records of 1-day prices of N=325 Iran stocks for the 5-yr period 2005-2010.

IV. STATISTICS OF CORRELATION COEFFICIENT

We analyze the distribution $P(C_{ij})$ of the elements $\{C_{ij}$; i#j$\}$ of the cross-correlation matrix C. We find this for 1-day returns from the database for the 5-yr periods 2005-2010 [Fig. 1]. We find that $P(C_{ij})$ has skewness and a positive mean value ($<C_{ij}>$ = 0.0104), implying that positively-correlated behavior is more prevalent than negatively-correlated (anti-correlated) behavior. We contrast $P(C_{ij})$ with a control —a correlation matrix R with elements R_{ij} constructed from N = 325 mutually-uncorrelated time series, each of length L = 1291, generated using the empirically found distribution of stock returns. Figure (1) shows that $P(R_{ij})$ is consistent with a Gaussian with zero mean. In addition, we see that the part of $P(C_{ij})$ for $C_{ij} < 0$ (which corresponds to anti-correlations) is within the Gaussian curve for the control, suggesting the possibility that the observed negative cross-correlations in C may be an effect of randomness [5]. The distribution shape is similar to the normal distribution and by comparing with Chinese and US stock markets [5,15], we find good agreement with findings of Chinese market that the correlation coefficients have normal distribution shape [15], but the US market does not have a normal behavior [5]. Both of the Chinese and Tehran stock markets are emerging markets. As an important note, TEPIX has a normal behavior same as Chinese market, but the mean value of cross-correlation of the stocks of the Tehran stock market is near to zero same as US market, and is smaller than Chinese market.

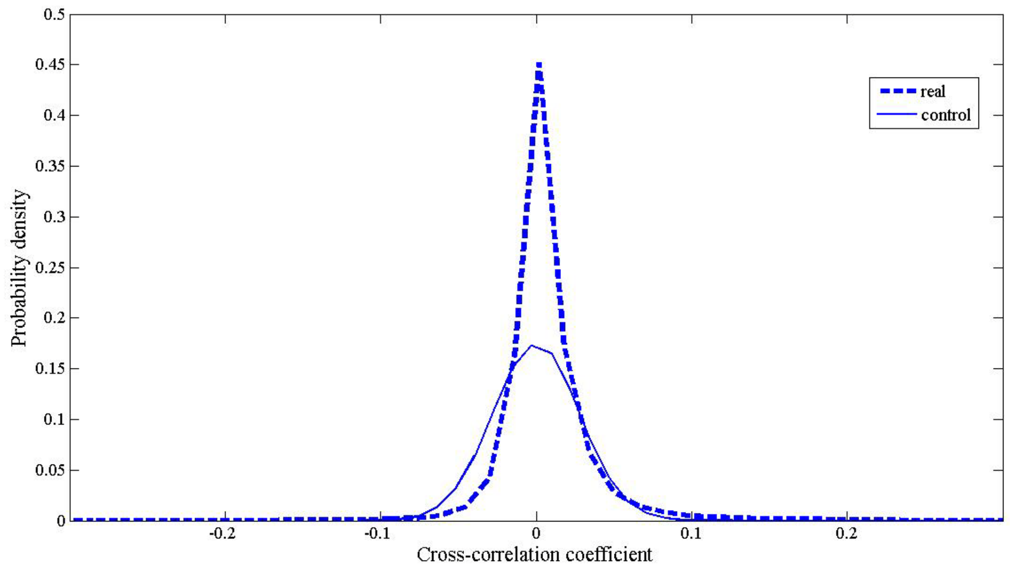

FIG. 1. $P(C_{ij})$ For C calculated using 1-day returns of 325 stocks for the period 2005-2010 (dashed curve). The solid curve shows the distribution of correlation coefficients for the control $P(R_{ij})$ of Eq. (5), which is consistent with a Gaussian distribution with zero mean.

V. EIGENVALUE DISTRIBUTION OF THE CORRELATION MATRIX

As stated above, we want to extract information from C. So, we compare the properties of C with those of a random cross-correlation matrix [16]. In the matrix notation, the correlation matrix can be expressed as

$$C = \frac{1}{L} G\, G^T \qquad (4)$$

Where G is an N×L matrix with elements $\{g_{it} : i=1,\ldots,N;\ t=1,\ldots,;L\}$

The spectral properties of C may be compared to those of a "random" Wishart correlation matrix. [5,17]

$$R = \frac{1}{L}\ A\, A^T \qquad (5)$$

Where A is an N×L matrix containing N time series of L random elements with zero mean and unit variance, that are mutually uncorrelated.

Statistical properties of random matrices such as R are known for many years in the physics literature [18] and have been applied to financial problems relatively recently [19].

In particular, the limiting property for the sample size N→ ∞ and sample length T→∞ , providing that $Q = \frac{T}{N} \geq 1$ is fixed, has been examined to show analytically that the distribution of eigenvalues λ of the random correlation matrix R is given by:

$$P_{rm}\,(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda_- - \lambda)}}{\lambda} \qquad (6)$$

for λ within the region $\lambda_- \leq \lambda \leq \lambda_+$, where λ_- and λ_+ are given by :

$$\lambda_\pm = \sigma(1 + \frac{1}{Q} + 2\sqrt{\frac{1}{Q}}\ ) \qquad (7)$$

Where σ is the standard deviation of the elements of G; (for G normalized this is equal to unity).

λ_+ and λ_- are the bounds of the theoretical eigenvalue distribution. Eigenvalues that are outside this region are said to deviate from Random Matrix Theory (RMT)[19]. First, we compare the eigenvalue distribution $P(\lambda)$ of C with P_{rm} (λ) [5]. Figure (2) show this comparison. We find the presence of a well-defined '"bulk" of eigenvalues which fall within the bounds [λ_- ,λ_+]for P_{rm} (λ) and for a few largest and smallest eigenvalues, there is deviations. Secondly, we contrast $P(\lambda)$ with the RMT result P_{rm} (λ) for the random correlation matrix of Eq. (5) , constructed from N= 325 mutually uncorrelated time series, each of the same length L=1291.we find good agreement with Eq.(6)(Fig.3). By comparison , we find good agreement with efficient market finding[5] .

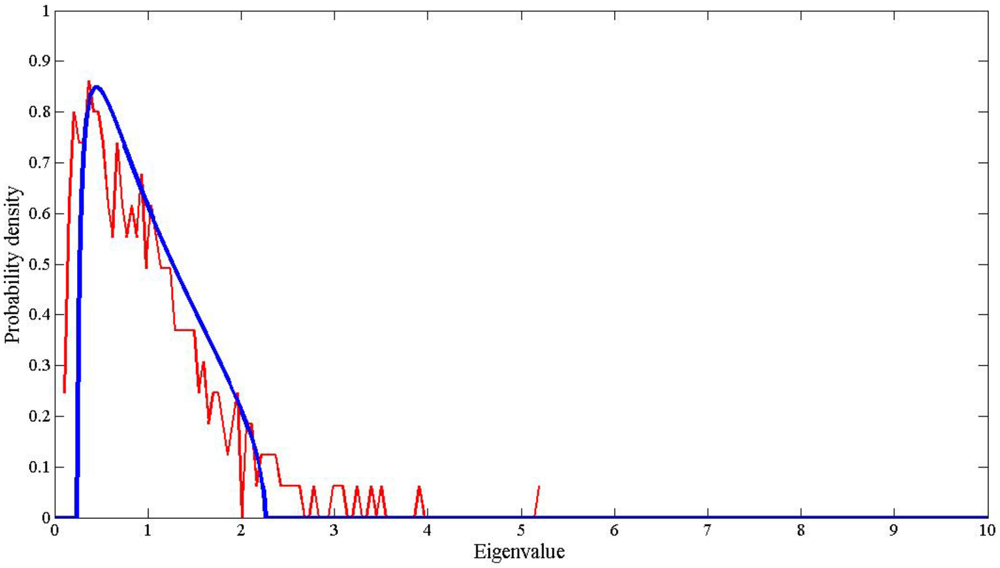

FIG. 2.Eigenvalue distribution P(λ) for C constructed from the 1-day returns for the stocks. The solid curve shows the RMT result $P_{rm}(\lambda)$ of Eq.(6). We see several eigenvalues outside the RMT upper bound λ_+ .

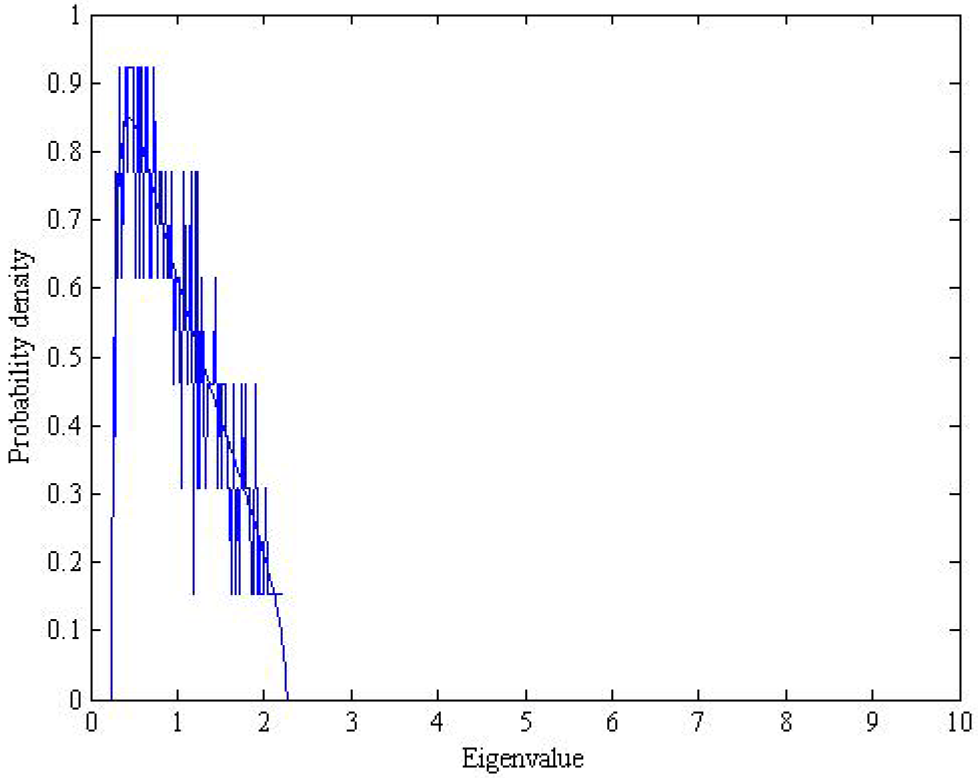

FIG. 3. Eigenvalue distribution for random correlation matrix, computed from N=325 computer-generated random uncorrelated time series with length L=1291 shows good agreement with the RMT findings.

VI. STATISTICS OF EIGENVECTORS

A. Distribution of eigenvector components

According to Fig (2), we can hope to see that these deviations should also be displayed in the statistics of the corresponding eigenvector components [16]. So, we analyze the distribution of eigenvector components. The distribution of the components of eigenvector u^k of a random correlation matrix R should conform to a Gaussian distribution with mean zero and unit variance.

We compare the distribution of eigenvector components of C with the Gaussian distribution (Fig.4). As we can see in this Figure, the distribution of the eigenvector components begin**s** from the first eigenvector as a quasi-normal plot to a non-normal curve at the last eigenvector and so, we can infer that, outside the bulk with increasing the eigenvalue, the normality behavior decreases. We analyze the eigenvector components of the eigenvalues belonging to the bulk and find good agreement with RMT result that show normal behavior and then analyze the eigenvectors corresponding to the largest and smallest eigenvalues and find good agreement with findings of Figure (4) too(Fig.5). We test the kurtosis of these eigenvectors by a Gaussian distribution that has a value 3 and find significant deviation from this distribution for some of the smallest and largest eigenvalues(Fig. (6)).Although, The eigenvector corresponding to the largest eigenvalue has kurtosis about 3 and skewness about zero, as same as the Gaussian distribution, this eigenvector has non-Gaussian behavior.

Finally, we infer that for eigenvalues smaller than λ_- there is a quasi-Gaussian behavior with high kurtosis and low and negative skewness, but, with increasing eigenvalue and arrival in bulk region, we can see a Gaussian behavior and then with exiting from the bulk region, we see non Gaussian behavior with intermittent kurtosis and skewness. The trend of these findings is consistent with mature market findings [5].

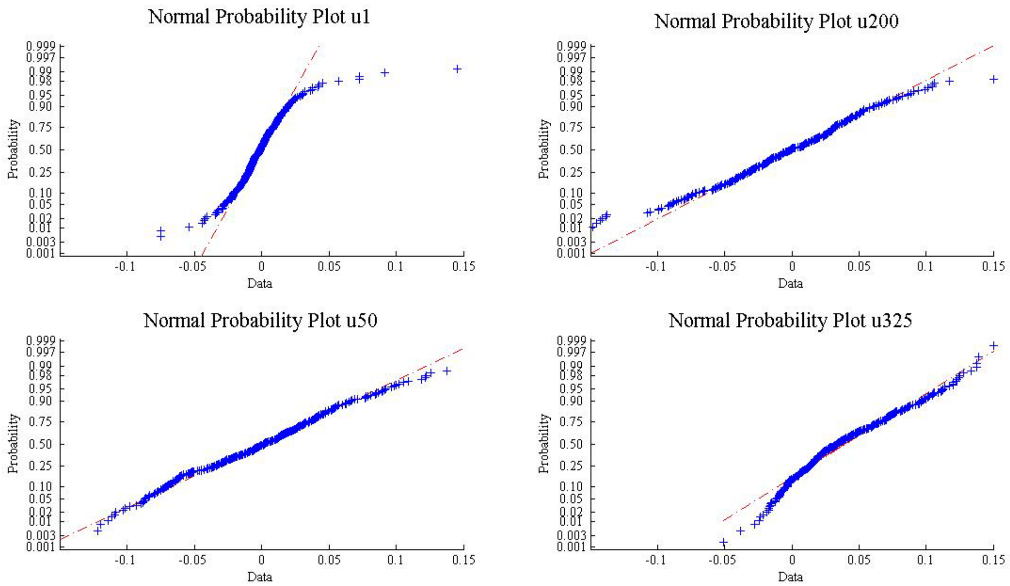

FIG. 4.normal probability plot for four eigenvectors. As we can see with incresing eigenvalue,normal behavior decreases.

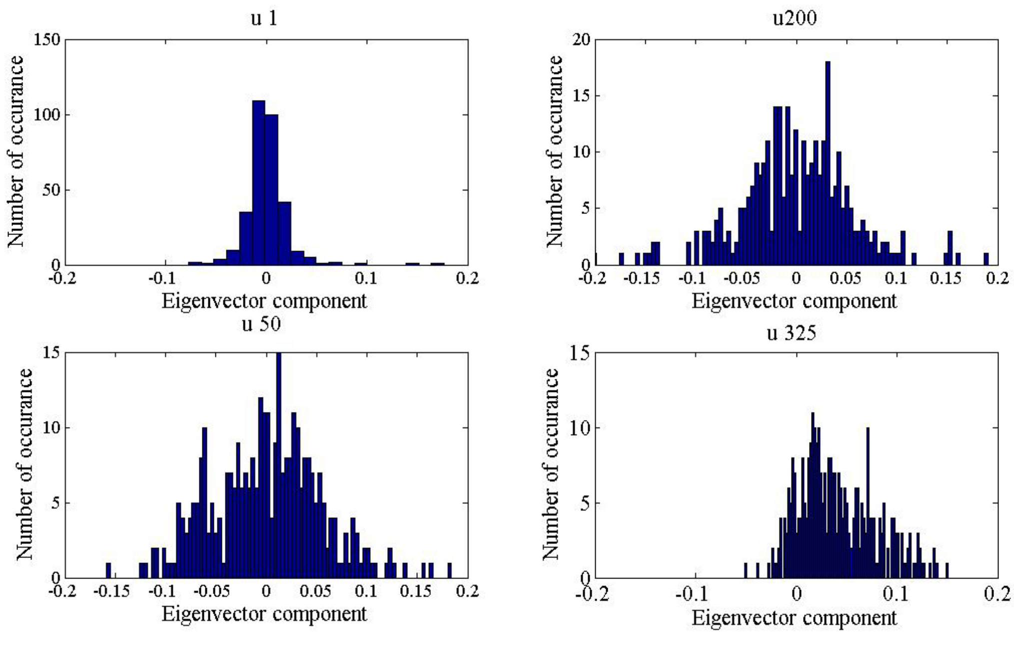

FIG. 5. Four eigenvector components, that shows for bulk(λ_{50},λ_{200}) there is Gaussian behavior but for eigenvalues larger than λ_{+} there is no Gaussian behavior. For eigenvectors corresponding to the smallest eigenvalues there is a quasi-Gaussian distribution.

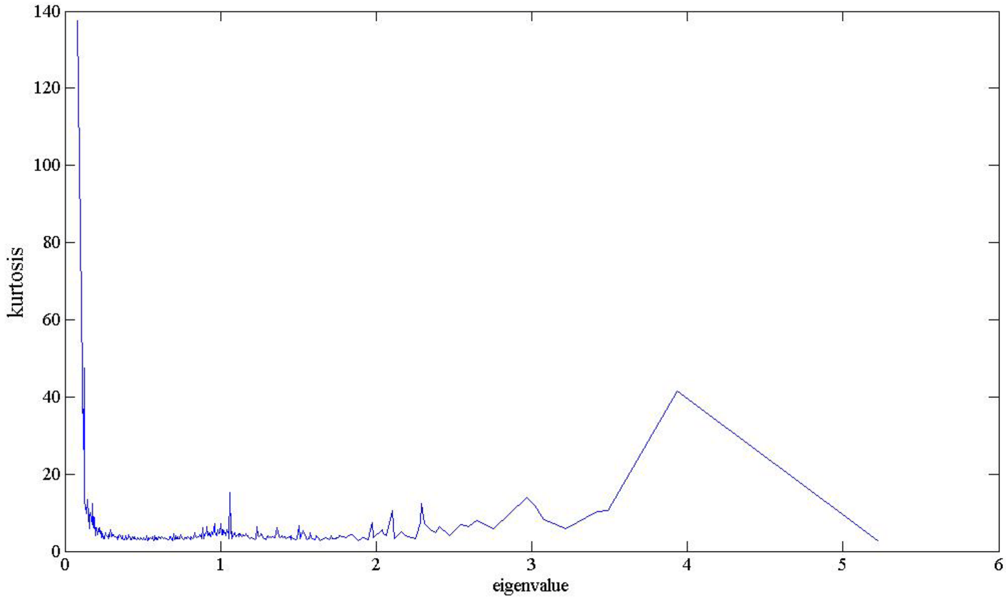

FIG. 6. Kurtosis of eigenvectors plotted against eigenvalues , and shows that largest eigenvalue has one of the least kurtosis And the smallest eigenvalues , have maximum eigenvector kurtosis. The bulk has a uniform-shaped kurtosis around 3 that is consistent with Gaussian distribution.

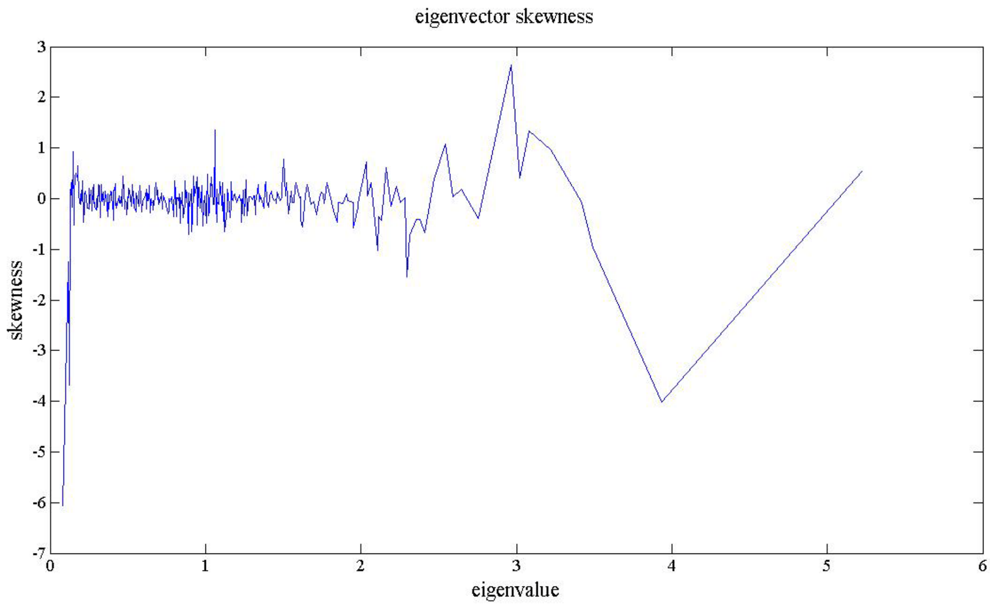

FIG. 7.Skewness of eigenvectors plotted against eigenvalue and shows that for smallest eigenvector, there is a negative and small skewness and for the largest eigenvalue, skewness is near to zero and for bulk there is a uniform-shaped skewness around zero that is consistent with Gaussian behavior.

B. Interpretation of the largest eigenvalue and the corresponding eigenvector

Since majority of the components, participate in the eigenvector corresponding to the largest eigenvalue, it can represent an influence that is common to whole market. Thus, the largest eigenvector quantifies the qualitative notion that certain newsbreaks (e.g., an interest rate increase) affect all stocks alike [4]. We investigate this notion by comparing the projection (scalar product) of the time series G(time series of returns) on the eigenvector U^{325}[5], with a standard measure of Tehran stock market performance — the returns of TEPIX index. We calculate the projection $G^{325}(t)$ of the time series Gj(t) on the eigenvector u^{325},

$$G^{325}(t) \equiv \sum_{j=1}^{325} u_j^{325} \quad Gj(t) \qquad (8)$$

By definition, $G^{325}(t)$ shows the return of the portfolio defined by u^{325}. We compare $G^{325}(t)$ with the returns of TEPIX index, and find large value of the correlation coefficient. Figure (8) shows $G^{325}(t)$, $G^{200}(t)$,$G^{50}(t)$ and $G^{1}(t)$ regressed against the returns of TEPIX index**.** Whenever the symmetry is high, we cannot see any good trend, but with increasing λ, we can see exiting from symmetry behavior and this means, there is information about our system and there is high correlation between components of u^{325} and market. We interpret the eigenvector u^{325} as quantifying market-wide influences on all stocks [20]. One way of statistically modeling an influence that is common to all stocks is to express the return G_i of stock i as

$G_i(t) = \alpha_i + \beta_i M(t) + \epsilon_i(t)$ (9)

where M(t) is an common term that is the same for all stocks, $<\epsilon (t)> = 0$, α_i and β_i are stock-specific constants, and $<M(t) \epsilon (t)> = 0$. The decomposition of Eq. (9) forms the basis of economic models, such as the Capital Asset Pricing Model and multi-factor models [4,5]. Since the components of the largest eigenvector (u^{325}) quantify market-wide influence on its all stocks, we

approximate M (t) with the G^{325} (t). The parameters α_i and β_i can therefore be estimated by an ordinary least squares regression.
Next, we remove the effect of u^{325} to each time series $G_i(t)$, and construct C from the residuals $\epsilon_i(t)$ of Eq. (9). Figure (9) shows the distribution $P(C_{ij})$. It has significantly smaller average value $<C_{ij}>$, showing that a large degree of cross-correlations contained in C is the effect of the largest eigenvalue [5].
By comparing with efficient market, we can see, there is high correlation between these $G^k(t)$ and TEPIX return for any eigenvalue, but in mature market, there is smaller correlation coefficients and for eigenvalue in the bulk there is no any linear relation[5]. For C in mature markets, decreasing trend after removal of market effect, in $<C_{ij}>$ is more than TEPIX's [5]. This may be because of importance of largest eigenvector and its components in mature markets that have more effects on the market but in TEPIX we see randomness in the majority of eigenvectors.

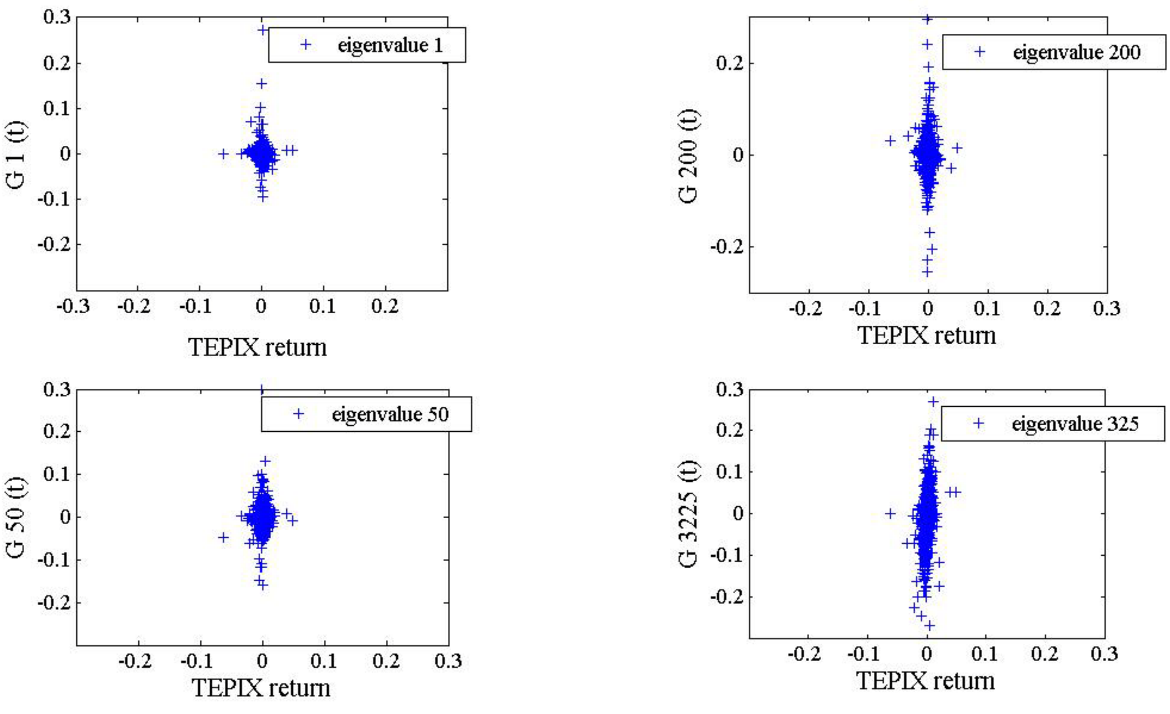

FIG. 8. Figure (8) shows $G^{325}(t)$, $G^{200}(t)$,$G^{50}(t)$ and $G^1(t)$ regressed against the returns of TEPIX index**.** Whenever the symmetry is high, we cannot see any good trend, but with increasing λ, we can see exiting from symmetry behavior and this means, there is information about our system and there is high correlation between components of u^{325} and market.

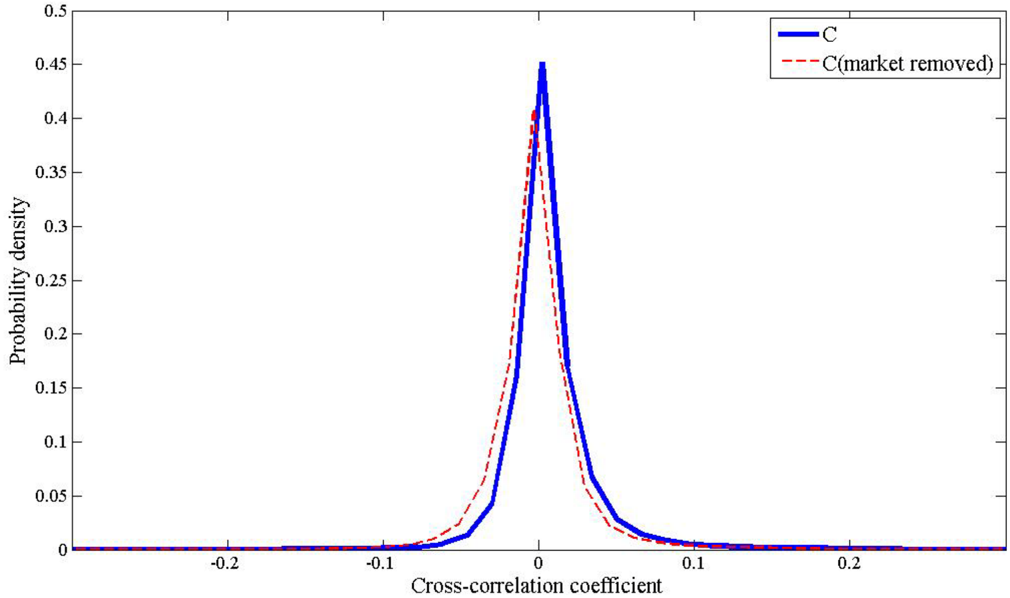

FIG. 9. $P(C_{ij})$ for stocks before and after removing the effect of market. <C_{ij}> before removing is .0104 and after removing the effect is .0026.

C. Number of significant participants in an eigenvector: Inverse Participation Ratio

After focusing on the largest eigenvalue which means market effect, we next focus on the remaining eigenvalues. Since proximity to the upper bound of bulk increases the effects of randomness, we use the notion of the inverse participation ratio (IPR) as the number of components that participate significantly in each eigenvector, which in turn reflects the degree of deviation from RMT result for the distribution of eigenvector components. The IPR of the eigenvector u^k is defined as

$$I^k \equiv \sum_{l=1}^{N} (u_l^k)^4 \qquad (10)$$

where u_l^k, l=1,…,325 are the components of eigenvector u^k. The meaning of I^k can be illustrated by two limiting cases: (i) a vector with identical components $u_l^k \equiv \frac{1}{\sqrt{N}}$ has $I^k = \frac{1}{N}$, whereas (ii) a vector with one component $u_l^k = 1$ and the remainder zero has I^k=1. Thus, the IPR quantifies the reciprocal of the number of eigenvector components that contribute significantly [5].

Figure 10(a) shows I^k for a computer-generated matrix of uncorrelated time series.

Figure 10(b) shows the IPR for C**.** This Figure shows us that for bulk region there is same behavior, but, on the edges of the eigenvalue spectrum of C we can see significant deviation from random control IPR. The largest eigenvalue has $1/I^k \approx 113$, showing that almost the majority of stocks participate in the largest eigenvector. For other deviating eigenvalues, there is different IPR depending on the number of significant eigenvector components.

In addition, we also find that there are large I^k values for eigenvectors corresponding to few of the eigenvalues smaller than λ_-. The deviations at both edges of the eigenvalue spectrum are considerably larger than mean of random control IPR, which suggests that the vectors are localized [5]—i.e only a few stocks contribute to them [5]. The presence of vectors with large values of I^k also arises in the theory of Anderson localization [5,20]. Whenever we compare this Figure with US market, we see, in the mature markets there is more difference between IPR of the largest eigenvalue and the other eigenvalues [5]. but in emerging market such an south Korean market[20], same as TEPIX, we can find this difference is small. This means that there are a few significant components in the maximum eigenvector or randomness in the market eigenvector is high.

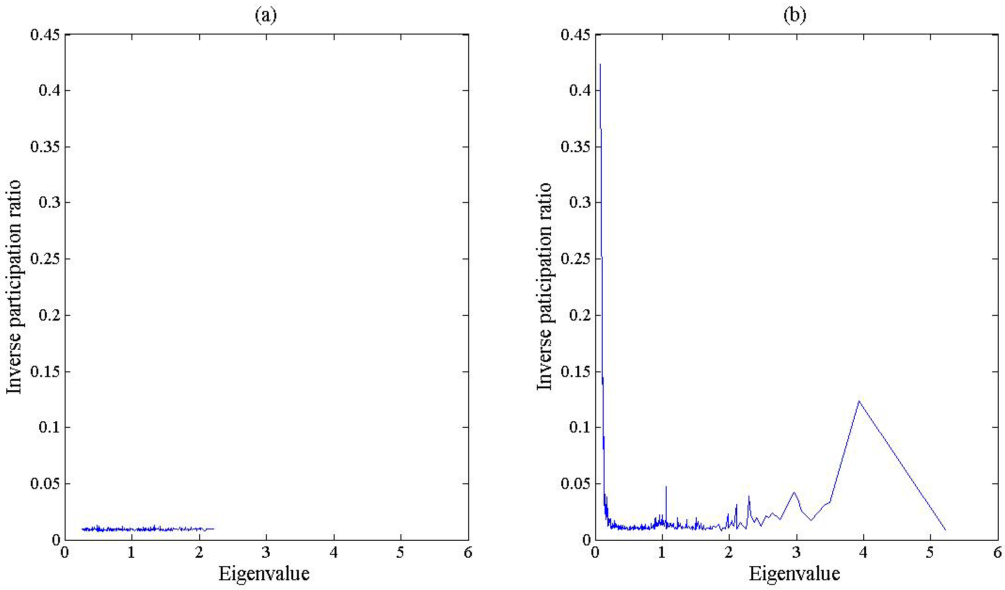

FIG. 10 (a) shows I^k for a computer-generated matrix of uncorrelated time series (b) shows the IPR for C**.** This Figure shows us that for bulk region there is same behavior, but, on the edges of the eigenvalue spectrum of C we can see significant deviation from random control IPR. We find this figure same as the figure (6) that shows the kurtosis of eigenvector components. That is not strange case because of their formulas'.

D. Interpretation of largest deviating eigenvectors :

In order to avoid the effect of λ_{325} , and thus $G^{325}(t)$, on the returns of each stock $G_i(t)$,we perform the regression of Eq.(9) and compute the residuals. we then calculate the correlation matrix C using these residuals . Next, we compute the eigenvectors of C thus obtained, and analyze their significant participants [5].

We find that each of these deviating eigenvectors contains stocks belonging to similar or related industries as significant contributors. We find that these eigenvectors partition the set of all stocks into distinct groups [21],which contain stocks of firms in the oil industry(u^{325}), mining sector and investment firms(u^{324}),drug industry and mining sector(u^{323}), agricultural sector(u^{322}),agricultural and mining sectors(u^{321}),agricultural and mining sectors(u^{320}), agricultural and mining sectors(u^{319}),car industry and agricultural sector(u^{318}), banking firms and mining sector(u^{317}) and mining sector and car industry(u^{316}).

It is interesting to know that majority of components of eigenvectors corresponding to the largest deviating eigenvalues are categorized in agricultural, mining sector and car industry or related businesses. It has good agreement with this fact that market of Iran has an agricultural and mineral base and the importance of the third major component are because of the government supports.

VII. Conclusion and Recommendations

We analyzed cross-correlation between return fluctuations of different stocks by using random matrix theory (RMT). We tested the statistics of eigenvalues of cross-correlation (C) between stocks of Tehran price index (TEPIX) as an emerging market in the period of 1 April 2005 to 1 April 2010 and compare these with a mature market (US market)[5]. We found that a majority of the eigenvalues of C fall within the bulk (RMT bounds between λ_- and λ_+) for the eigenvalues of the random correlation matrices. Further, we found that the distribution of eigenvector components for the eigenvectors corresponding to the largest deviating eigenvalues, display systematic deviations from the RMT prediction. We analyzed the components of the deviating eigenvectors and found that the largest eigenvalue corresponds to an influence common to all stocks. Our analysis of the remaining deviating eigenvectors shows distinct sectors, whose identities corresponds to the structure of the market. Market of Iran is a government- based market. We compare some of these attributes with US, Chinese and Korean markets [15, 20] and found good agreement with Chinese and Korean markets but some crucial difference between US and Iran markets. For future researches, we propose to investigate more about these relationships from more views and for more countries.

References:

[1] J. D. Farmer, Comput. Sci. Eng. 1, 26 (1999).
[2] R. N. Mantegna and H. E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance (Cambridge University Press, Cambridge 1999).
[3] J. P. Bouchaud and M. Potters, Theory of Financial Risk (Cambridge University Press, Cambridge 2000).
[4] J. Campbell, A.W. Lo, A. C. MacKinlay, The Econometrics of Financial Markets (Princeton University Press, Princeton, 1997)

[5]Plerou, V.,Gopikrishnan, P.,Rosenow, B.,Amaral, L.,Guhr, T., Stanley, H.E., Arandom matrix approach to cross-correlation in financial data, Phys.Rev.E65 (2002) 066126.

[6] E. P. Wigner, Ann. Math. 53, 36 (1951); Proc. Cambridge Philos. Soc. 47 790 (1951).
[7] E. P. Wigner, "Results and theory of resonance absorption," in Conference on Neutron Physics by Time-offlight (Oak Ridge National Laboratories Press, Gatlinburg, Tennessee, 1956), pp. 59.
[8] E. P. Wigner, Proc. Cambridge Philos. Soc. 47, 790 (1951).
[9] F. J. Dyson, J. Math. Phys. 3, 140 (1962).
[10]F. J. Dyson and M. L. Mehta, J. Math. Phys. 4, 701 (1963)..

[11] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, Phys. Rev. Lett. 83, 1469 (1999).
[12] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, Phys. Rev. Lett. 83, 1471 (1999)

[13] rahavardnovin.com

[14] Only those stocks, which have survived the 5-yr period 2005-2010 and traded more than 200 days, were considered in our analysis.

[15] Wei.Qiang, Hung,., .Xin-Tian ,Zhuang.,Shuang,Yao., A network analysis of the Chinese stock market, Physica A 388(2009) 2956-2964.

[16] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, Phys. Rev. Lett. 83, 1469 (1999).

[17]Laloux,L.,Cizeau,P.,Potters,M.,Bouchaud,J.,Random matrix theory and financial correlations,Int.J.Theoret.&Appl.Finance 3(3)(2000) 391-397.

[18]Mehta,M.L.,Random Matrices,Third Edition,Elsevier,2004.

[19]Conlon,T.,Ruskin,H.J.,Crane,M.,Random Matrix Theory and Fund of Funds Portfolio Optimization,q-fin.ST,2010.

[20]Gyuchang, Lim.,Soo Yong, Kim.,Junghwan, Kim., Pyungsoo,Kim.,Yoonjong, Kang., Sanghoon,Park.,Inho,Park.,Sang-Bum,Park.,Kyungsik,Kim., Structure of a financial cross-correlation under attack, Physica A 388(2009) 3851-3858

[21]Names of the stocks in the largest deviating eigenvectors are available upon a request.
[22] in this paper, we suppose that mature market is efficient market.