AN ANALYSIS OF EIGENVECTORS OF A STOCK MARKET CROSS-CORRELATION MATRIX

Hieu Nguyen   Phuong Tran   Quang Nguyen

John von Neumann Institute

Vietnam National University - Ho Chi Minh City

***Abstract*—Random matrix theory (RMT) has been used to analyze the structure of the stock returns cross-correlation matrix. Those authors found common results in various markets: only a few eigen-mode (portfolio) appear significant and the rest resemble that of a random matrix. In concrete, there is a very large eigenvalue representing the whole market mode followed by several other large ones, plus the bulk which mostly agrees with the theoretical spectrum distribution predicted by the RMT theory. However, there are not many studies focusing on the eigenvector components of such modes. In this work, we analyze in detail the components of the eigenvectors and found that components of the market eigenvector depends linearly to the degree of correlation of each stock. Therefore, the corresponding market portfolio must be considered as a market-correlation-portfolio rather than the market-cap-portfolio. Other informative eigenvectors also show important structures. Those results could be very meaningful when one analyze the structure of the financial market and their applications.**

Keywords—Random Matrix Theory, Cross-Correlation

I. INTRODUCTION

The study of the cross-correlation between stocks is of great importance not only for understanding the complex structure of the stock market, but also for the asset allocation and risk management. Several established statistical methods employed from physics have been used to help studying the cross-correlation matrix including the minimal spanning tree (MST) (first applied in [1]) and random matrix theory (RMT) (first applied in [2], [3]). The MST method construct a tree that connect all stocks in the market from which one can analyse the market topology, which is usually represented by sector. On the other hand, the RMT method is powerful noise filter that reveals important correlation modes in the market.

Under the lenses of RMT, the eigenvalues and eigenvectors of the cross-correlation matrix were calculated and compared to those of a random matrix, generated from randomly generated uncorrelated time-series (hereby call RMT eigenvalues and eigenvectors). It was found in common literature [2]–[12] that the majority of eigenvalues from a stock returns cross-correlation matrix overlap with the bulk of RMT eigenvalues, except:

- The largest eigenvalue λ_1
- Several large eigenvalues higher than the maximum theoretical eigenvalue λ_+ (to be commented further on)
- Some very small eigenvalues smaller than the minimum theoretical eigenvalue λ_-

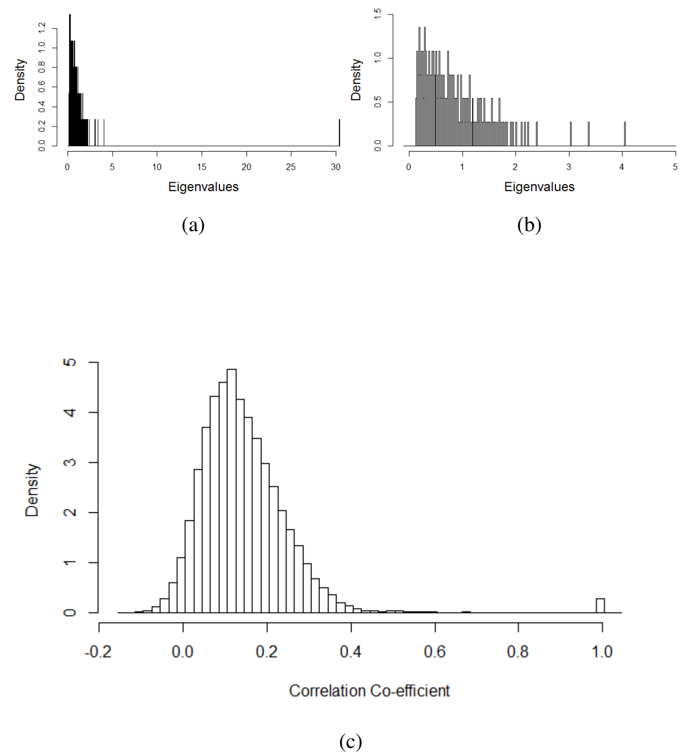


Fig. 1: (a): Distribution of all empirical eigenvalues (b): Distribution of all empirical eigenvalues without the largest eigenvalue (c): Distribution of Correlation Coefficients of the Correlation Matrix

One can easily see the largest eigenvalue, hereby called λ_1, is much higher than subsequent eigenvalues. This implies that most stocks follow a dominant correlation mode: there is a tendency of a positive mutual correlation of all stocks in the market. From the financial point of view, it is intuitive to understand because all listed companies operate in the same economy and therefore, are influenced by many (macro) common factors. This view is also coherent with the CAPM theory and the co-movement of the market as a whole is called the market mode.

From Fig. 1(c) it is seen the correlation coefficients are distributed in a positive narrow range and mostly positive. In our previous work [13] we modeled the correlation matrix by a constant correlation ρ_1 (one-factor correlation matrix model) and found the appearance of the largest eigenvalue

λ_1. Furthermore, the value of λ_1 can be approximated by the product of ρ_1 and N, where N is the total number of stocks. While there are ample studies focusing on dynamics of the eigenvalues, fewer works are found about the eigenvectors. Most studies suggest that the eigenvector component of the largest eigenvalue indicates a uniform distribution of weights among stocks. However, we re-examined this distribution (Fig. 3(a)) and attempt to relate it to the correlation degree of each stock. We define correlation weight by:

$$w_i = \sum_{j=1}^{N} (\rho_{ij}) \tag{1}$$

and found that a_i^1, the component i of the eigenvector correspond to the largest eigenvalue λ_1, is linearly dependent on w_i, as demonstrated in Fig. 3(b). To our knowledge this result is not addressed yet in studies using RMT for financial markets. It is straightforward to interpret the portfolio associated to this eigenvector to be a correlation-weight-portfolio rather than other definition of market-portfolio (which is usually weighted by stock market capitalisation or a market-uniform portfolio). For example, in the Vietnamese market, the stock that is most weighted in this market mode is SSI, which is a brokerage firm which is a mid-cap. The risk and behaviour of such portfolio is a subject for our future work.
We also study the components of eigenvectors associated with other large eigenvalues smaller than the top eigenvalue but lying outside of the bulk predicted by RMT. Considering that the components of the market-correlation-portfolio are mostly negative except for a few positive ones, the components of other eigenvector are expected to be distributed around zero in order to satisfy the orthogonality condition.
In addition, eigenvectors with significant positive and negative components values may indicate groups of stocks belonging to specific financial sectors. Other studies in developed markets such as the US, Japan found that they represent industry sector (see [3]–[5]) while some studies in emerging markets such as China and India found a mix of industry sectors and other specific treatment stock groups, as in [9]–[12]. In the Vietnamese stock market, we show that one of the major groups is the speculative group. In fact, it depends on the property of the market and their investors behaviour that create such structure. We also found that within each mode, the *long* and the *short* groups need to be positively correlated within each group and possibly negatively correlated between groups.
The paper is organised as follows: section 2 presents the data and method, section 3 presents our main results and in section 4 we discuss, conclude and suggest some future work.

II. Data and Methods

A. Data

The data is collected from the two Vietnamese stock exchanges - the Hanoi Stock Exchange and the Ho Chi Minh Stock Exchange - for the period from 1/1/2015 to 31/5/2017. We selected active stocks that were listed prior to 2015. Moreover, to ensure some degree of liquidity, the stocks must have, on daily average, trading volume exceeding 50,000 units and transaction value exceeding one billion Vietnam Dong. The prices on days with no trades, due to technical reasons such as a change of stock exchange, are assumed to be constant with the last trading day. In the end, our dataset consists of N = 186 stocks with daily prices spanning over a total of T = 600 days.
We consider the standardized stock returns as follows: (i) we first obtain the log-returns $R_i(t)$ from the daily closing prices $R_i(t) = log(S_i(t+1)) - log(S_i(t))$ (ii) we then remove the (empirical) mean of each stock from its returns: $R_i^*(t) = R_i(t) - \langle R_i \rangle$ where $\langle . \rangle$ denotes the sample average; (iii) we finally normalize each stocks returns by its volatility which, in this work, is simply estimated by the standard deviation: $r_i(t) = \frac{R_i^*(t)}{\sigma(R_i)}$. From this, we obtain an estimation of the $N \times T$ correlation matrix C of returns using the Pearson estimator for the correlation co-efficients C_{ij}:

$$C_{ij} = \langle r_i(t) r_j(t) \rangle$$

B. Methods

Once the eigenvalues and (normalised) eigenvectors are obtained from the estimated correlation matrix E, the eigenvalues are plotted and compared against a theoretical eigenvalues distribution given by Random Matrix Theory. Deviations of the empirical eigenvalues from the theoretical distribution then suggests the presence of true informative eigenvalues, separated from noise-corrupted eigenvalues.
In a more general setting, let N and T be respectively the number of stocks and the number of realizations of data for each stock. We consider a random matrix called the Wilshart matrix of the form $W = \frac{1}{T} Y Y^*$ where Y is a NxT matrix of T i.i.d vectors of size N. We note that in this context, the Wilshart matrix is in fact the sample covariance matrix (or correlation matrix, after normalization). From the classic paper [14], it is known that: in the large limits of T and N to infinity such that $Q := \frac{T}{N} > 1$ and under the null hypothesis of a purely random matrix, the limiting spectral density - more famously known as the *Marchenko-Pastur* (MP) density - reads:

$$\rho_{MP}(\lambda) = \frac{Q}{2\pi\sigma^2} \cdot \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

where λ_+, λ_- are the upper and lower bounds of the bulk component containing all eigenvalues:

$$\lambda_\pm = \sigma^2 \cdot (1 \pm \frac{1}{\sqrt{Q}})^2$$

and σ^2 is equal to the variance of the elements, which is 1 given the normalisation of stock returns.
In order to highlight the non-randomness in the components of informative eigenvectors, as in [2], we may also compare these components (normalized such that the eigenvector has length $\sqrt{N}$) against the entropy-maximizing Porter-Thomas distribution - in this case simply the standard normal distribution:

$$P(u) = \frac{1}{\sqrt{2\pi}} \cdot exp(\frac{-u^2}{2})$$

If there is no information in an eigenvector, we expect the components to obey the distribution.

III. RESULTS

A. Overview

As expected, there is one outstanding eigenvalue at $\lambda_1 = 30.4$ well separated from the rest, a bulk containing the majority of eigenvalues and a few others outside this bulk but lower in value than the largest one.

Fig1(a) shows the correlation coefficients distributed in a narrow range, indicating that there exists a coherent movement of all the stocks 2013 the market mode.

In order to separate the informative from the noisy, we first compare the empirical eigenvalues distribution with the MP-density where $Q = \frac{T}{N} = \frac{599}{186} \approx 3.22$ and $\sigma^2 = 1$ as per normalization.

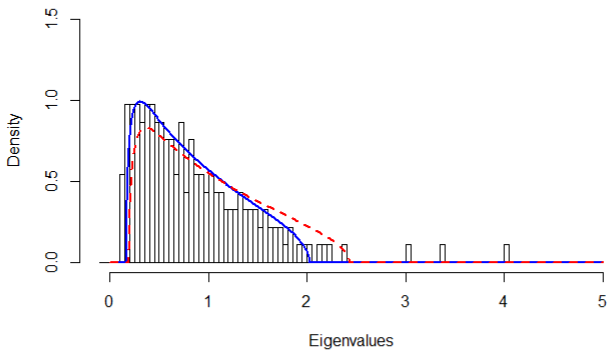


Fig. 2: Distribution of the eigenvalues, excluding the top eigenvalue. Also included is a plot of the MP density with $\sigma^2 = 1$ (dotted line) and with $\sigma^2 = 0.836$ (solid line)

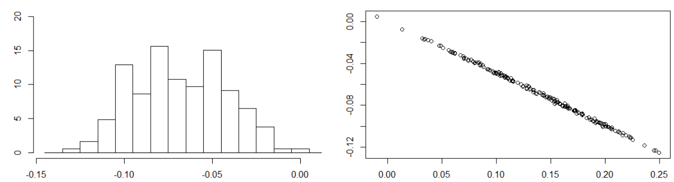
As in the work of Laloux et. al. [3], a fit of the MP - density can also be constructed with a modified $\sigma^2 = 1 - \frac{\lambda_1}{N} = 0.836$. This corresponds to the observation that the top eigenvalue and eigenvector is clearly not random and so it is reasonable to subtract the contribution of the market eigenvalue from the variance of the random part. A fit of the corresponding plot is also included in Fig. 2.

We note that a better fit can be made, considering three reasons:(i) larger eigenvalues are prone to overestimation (ii) outlying eigenvalues containing true information may effectively reduce the variance σ^2 of the random part of the matrix (iii) volatility correlations may cause Q to deviate from the effective Q for the MP density.

Still the current fit is quite convincing, with 90% of the empirical eigenvalues lying within the bulk. In particular, the eigenvalues 1 to 4 (in order of magnitude) are likely to contain information.

B. On the Top Eigenvector

We continue with a study of the top eigenvector associated with the largest eigenvalue, often referred to as the market eigenvector, market mode or the market portfolio. Fig. 3 shows the components of the top eigenvector, where all the components are negative (or of the same sign). In fact, the sign of those components are not really important as a set of eigenvectors remain eigenvectors if we change the sign of all components of all eigenvectors. Therefore, we consider that the first eigenvector portfolio is a long-only portfolio. Fig. 3(b) shows the linear relationship between the components and the correlation weights defined in 1. This indicates a correlation-based structure to the market eigenvector, instead of a market capitalization-based weighting or a uniform weighting structure. Indeed, for our data, the stocks holding the most weight include the tickers SSI, HCM, VND, PVT etc. which are not the highest-cap stocks or, for some, not even high-caps in the market. Whether this structure is stable or can be proved theoretically is a subject for future work.

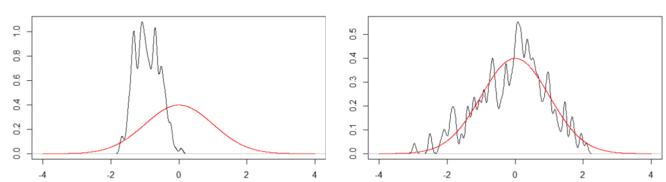


(a) (b)

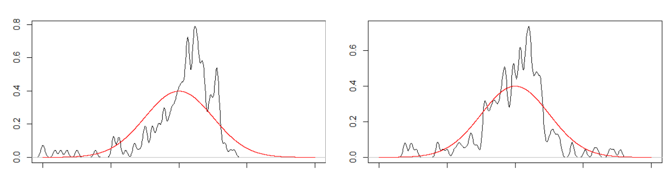
Fig. 3: (a): Distribution of market mode components (b): Plot of market mode components against corresponding correlation weights of the stocks

C. On subsequent Eigenvectors

We make the first observation that: since the market portfolio is long-only, to ensure orthogonality, all others eigenvectors must have both negative and positive components. Say differently, all other portfolios are some sort of long-short portfolio, as illustrated in Fig. (4), referred to from here on as *long* and *short sub-portfolios*.



(a) (b)



(c) (d)

Fig. 4: Smoothed densities of the components of the eigenvectors 1,2,3 and 4 (plot a,b,c and d respectively) against the Porter-Thomas distribution, plotted with a red thick line.

It has been documented in literature the *localization effect* displayed in the eigenvector components, that is, each eigenvector is dominated by a group of stocks,. This is particularly evident for developed markets [4] where the eigenvectors exhibit strong presence of financial sectors: the blue-chips group, the technology sector, the gold sector, to name a few. The case with emerging markets is less apparent where the top eigenvectors are dominated by fewer and less traditional sectors: for China (see [9], [10]), the sectors include blue-chip companies, the so-called ST stocks (the stocks with Special Treatment defined in [9]) and Shanghai-listed real estate businesses; still the eigenvectors exhibit localization in the component values. Fig. (3) however shows that for the Vietnamese stock market, this localization effect is not as clear-cut among different eigenvectors. Specifically, except the weights of the eigen-portfolio 4 and of the short-only sub-eigen-portfolio 3 where there are groups of dominating stocks, the weights are more spread out.

We shall look at the physical components of these eigenvectors to further understand this phenomenon. For each eigenvector, we consider the top 12 contributing stocks in each of the long and short sub-portfolios. The stocks are listed in Table I, together with a general description:

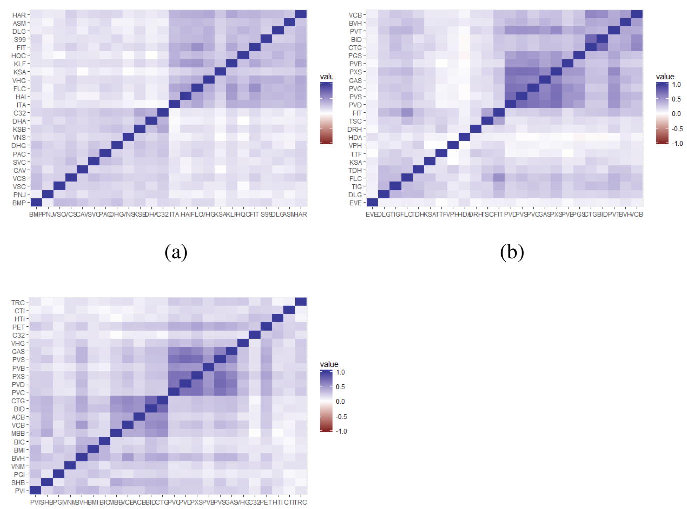
Eigen-Portfolio	Long Sub-Portfolio	Short Sub-Portfolio
Eigen 1	BMP, PNJ, VSC, VCS, CAV, SVC, PAC, DHG, VNS, KSB, DHA, C32	ITA, HAI, FLC, VHG, KSA, KLF, HQC, FIT, S99, DLG, ASM, HAR
	Mid and low-cap, strong upward trend overall, high EPS and ROE, low PE.	Downward trend, slump in profits, EPS < 1 while still maintaining solid liquidity
Eigen 2	EVE, DLG, TIG, FLC, TDH, KSA, TTF, VPH, HAD, DRH, TSC, FIT	PVD, PVS, PVS, GAS, PXS, PVB, PGS, CTG, BID, PVT, BVH, VCB
	Speculative mid-cap, on a downward trend, despite high income levels for some of the stocks.	Stocks in the Gas-Oil Sector and in the Banking Sector
Eigen 3	PVI, SHB, PGI, VNM, BVH, BMI, BIC, MBB, VCB, ACB, BID, CTG	PVC, PVD, PXS, PVB, PVS, GAS, VHG, C32, PET, HTI, CTI, TRC
	VNM, the highest-cap stock in Vietnam, and others from the Banking Sector.	Stocks in the Gas-Oil Sector

TABLE I: Notable physical components of the Eigen-portfolios 2,3 and 4

Curiously, the second eigen-portfolio is composed of mid-cap stocks whose performance has significantly slumped (forming the short sub-portfolio) or improved (forming the long sub-portfolio). That the portfolio is composed of mid and low-caps instead of blue-chip stocks (as in developed markets and to a lesser extent, emerging markets) is surprising. A similar observation can be made for the long sub-portfolio yielded by the third eigenvector, though the contributing stocks are of a different nature: they are mostly mid-cap and speculative with volatile price trajectories. It should be pointed out that these stocks are not entirely similar to the "ST" stocks defined in [9], which are stocks with an abnormal financial situation, with profits sharply decreasing over the data period.

The above observation leads to an interpretation that the eigenvectors with significant eigenvalues do not (only) reflect the dominating sectors within the economy, rather the investing appetites of the stock market participants. This view is particularly understandable for Vietnam where the presence of speculative individual investors, informed or not, is significant relative to that of financial institutions. It would be interesting to see whether for a shorter time-period, perhaps with intra-day data, it is possible to study the short-term investing behaviour of the market.

Meanwhile, the subsequent eigenvectors displays a more sectoral structure often seen in more developed markets, where the two largest sectors in Vietnam, Oil & Gas and Banking, dominate. Considering that few other sectors in Vietnam boast enough listed companies to form a meaningful group, the modest sectoral behaviour is expected; still, it is surprising the rather unsubstantial presence of the blue-chip stocks as a group within the eigen-portfolios.



(a) (b) (c)

Fig. 5: Correlation between the top contributing stocks of the long and short sub-portfolios within the eigen-portfolio 2, 3 and 4 (picture a, b and c respectively)

Finally, we observe the correlation structure of the eigen-portfolios. As indicated in the Fig. 5, there is strong correlation within each of the sub-portfolios and weak correlation between sub-portfolios.

IV. Discussions and Conclusion

In this work, we have analysed in details the components of the important eigenvectors of a cross-correlation matrix in the Vietnamese stock market. We found the following results: firstly, the component i of the eigenvector correspond to the largest eigenvalue λ_1 is linearly dependent on the correlation weight w_i of each stock. This finding is important because as λ_1 is far higher than other eigenvalue, its corresponding portfolio has a great contribution of variance (or risk) to any suitably diversified portfolio. We redefined this portfolio as a correlation-weighted market portfolio and will study its role in future work, as well as the relation above. Secondly, as found in literature, all markets have this common correlation-weighted market mode, we suppose that all stocks in the market have a tendency to be mutually positively correlated and the correlation weight w_i of each stock is positive. In consequence, the components of the eigenvector correspond to the largest eigenvalue λ_1 are positive and that of other eigenvectors are equally distributed around zero. In other words, if the correlation-weighted market portfolio is the common market mode, all other portfolios (corresponding to other eigenvectors) are kind of ”long-short” portfolios. Thirdly, we also highlight some important groups of contributing stocks in some modes and found mixed results on the nature of these groups compared to previous works on developed and emerging markets. In conclusion, this work focuses on the eigenvector component of cross-correlation matrix using RMT method. We have found some results that can be used in other risk management study, especially for a diversified portfolio. Further work is needed to complement these initial findings and get a better view about the stock market topology.

ACKNOWLEDGMENT

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number B2017-42-01.

REFERENCES

[1] Mantegna R.N., Hierarchical structure in financial markets, 1999. Eur. Phys. J. B 11, 193.

[2] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Physical review letters 83 (7) (1999) 1467.

[3] L. Laloux, P. Cizeau, M. Potters, J.-P. Bouchaud, Random matrix theory and financial correlations, International Journal of Theoretical and Applied Finance 3 (03) (2000) 3912013397.

[4] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, H. E. Stanley, Random matrix approach to cross correlations in financial data, Physical Review E 65 (6) (2002) 066126.

[5] A. Utsugi, K. Ino, M. Oshikawa, Random matrix theory analysis of cross correlations in financial markets, Phys. Rev. E 70 (2004) 026110

[6] D. Wilcox, T. Gebbie, On the analysis of cross-correlations in South African market data, Physica A 344 (2004) 294.

[7] G. Oh, C. Eom, F Wang, W.-S. Jung, H. E. Stanley, S. Kim, Statistical properties of cross-correlation in the Korean stock marketEur. Phys. J, B 79, (2011), 55.

[8] S. Cukur, M. Eryigit, R. Eryigit, Cross-Correlation in an Emerging Market Financial Data, Physics A 376 (2007) 555.

[9] Shen, J., Zheng, B. Cross-correlation in financial dynamics. Europhys. Lett. 86,48005 (2009).

[10] Jiang, X. F. Zheng, B. Anti-correlation and subsector structure in financial systems. EPL 97, 48006 (2012).

[11] R.K. Pan, S. Sinha, Collective behavior of stock price movements in an emerging market, Phys. Rev. E76, 046116 (2007)

[12] V. Kulkarni, N. Deo, Volatility of an Indian stock market : A random matrix approach, e-print (2005) arXiv:physics/0512169.

[13] Q.Nguyen, One-factor model for cross-correlation matrix in the Vietnamese stock market,Physica A 392(2013) 291520132923.

[14] V. A. Marchenko, L. A. Pastur, Distribution of eigenvalues for some sets of random matrices, Matematicheskii Sbornik 114 (4) (1967) 5072013536.