Abstract. The characteristic of Indonesian stock market is interesting especially because it represents developing countries. We investigate the dynamics and structures by using Random Matrix Theory (RMT). Here, we analyze the cross-correlation of the fluctuations of the daily closing price of stocks from the Indonesian Stock Exchange (IDX) between January 1, 2007, and October 28, 2014. The eigenvalue distribution of the correlation matrix consists of noise which is filtered out using the random matrix as a control. The bulk of the eigenvalue distribution conforms to the random matrix, allowing the separation of random noise from original data which is the deviating eigenvalues. From the deviating eigenvalues and the corresponding eigenvectors, we identify the intrinsic normal modes of the system and interpret their meaning based on qualitative and quantitative approach. The results show that the largest eigenvector represents the market-wide effect which has a predominantly common influence toward all stocks. The other eigenvectors represent highly correlated groups within the system. Furthermore, identification of the largest components of the eigenvectors shows the sector or background of the correlated groups. Interestingly, the result shows that there are mainly two clusters within IDX, natural and non-natural resource companies. We then decompose the correlation matrix to investigate the contribution of the correlated groups to the total correlation, and we find that IDX is still driven mainly by the market-wide effect.

1. Introduction

1.1. Motivation

Financial markets are complex systems in which a large number of traders interact with one another and react to external information in order to determine the best price for a given asset – animals, ores, equities, currencies, bonds, or their derivatives. In recent years, the study of stock markets has attracted much interest of physicists not only for practical purposes such as asset allocation and portfolio risk estimation but also for scientific reasons of understanding the underlying dynamics of complex systems. Relevant information about a given asset in a market is incorporated in the time series of its price. As stated in the efficient market hypothesis, all available information is instantly processed when it reaches the market and is reflected in a new value of prices of the assets. Therefore, the dynamics of markets can be studied through the fluctuations of the prices of its constituent assets.

The underlying structure of a complex system manifests itself in correlations among its constituents. However, analyses solely based on correlations are problematic mainly for two reasons: (i) market conditions vary with time and the cross correlations that exist between any
pair of stocks may not be stationary, and (ii) finite length of time series introduces measurement noise.

In this research, random matrix will be applied to distinguish the random and non-random property of the correlations. From the non-random part of the correlations, we want to identify the various modes within the market system which supposedly describe dominant correlated groups and also quantify the effects of these modes within the system.

1.2. Correlation Matrix

To construct the correlation matrix, we first calculate the return $G_i(t)$ of stock $i = 1, \ldots, N$ from the stock price time series $S_i(t)$, recorded over an interval time $\Delta t$,

$$G_i(t) \equiv \ln S_i(t + \Delta t) - \ln S_i(t).$$

(1)

The natural logarithmic approach is a scaling technique intended to produce a data set with a Gaussian distribution. Since different stocks have varying degree of volatility (standard deviation), we define the normalized return

$$g_i(t) \equiv \frac{G_i(t) - \langle G_i(t) \rangle}{\sigma_i},$$

(2)

where $\sigma_i \equiv \sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}$ is the standard deviation of $G_i$, and $\langle ... \rangle$ denotes time average over the period studied. We then calculate the correlation matrix $C$ with elements

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

(3)

or in matrix representation,

$$C = \frac{1}{L} GG^T,$$

(4)

where $G$ is an $N \times L$ matrix with elements $\{g_i(m\Delta t); i = 1, \ldots, N; m = 0, \ldots, L - 1\}$, whereas $G^T$ denotes the transpose of $G$.

The correlation matrix $C$ is an $N \times N$ matrix with each element $C_{ij}$ indicating the linear dependence between stocks $i$ and $j$. The value of the coefficient is within the range $-1 \leq C_{ij} \leq 1$, where $C_{ij} = 1$ implies perfect correlation, $C_{ij} = -1$ implies perfect anti-correlation, and $C_{ij} = 0$ implies no correlation between the pair of stocks.

From the correlation matrix, we then calculate the eigenvalues, $\lambda_i$, and the eigenvectors, $u^i$, to be compared with that of a similar correlation matrix constructed from a randomly generated matrix (random matrix) with the same size as $S_i(t)$.

1.3. Random Matrix Theory

Random matrix theory is basically the application of a matrix-valued random variable, called a random matrix, in various problems. In recent years, researches have shown that the spectral properties of a random matrix show high resemblance with that of a stock price data, indicating a high degree of randomness existing within the stock market. This notion is used to filter out the random effects from the stock market data by comparing the properties of the real data and the random matrix and distinguishing the parts of the real data which conforms with the random matrix as noise and those that deviate from the random matrix as genuine information.

Consider a random matrix $A$ with size $N \times L$ composed of random elements $a_{mn}$ with zero mean and unit variance. Its correlation matrix is calculated using Equation 4 and denoted $R$,

$$R = \frac{1}{L} AA^T.$$
The eigenvalue distribution $P_{rm}$ of the random correlation matrix $R$ has been shown analytically to follow

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

in the limit of $N, L \to \infty$ and $Q \equiv \frac{L}{N} \geq 1$ for $\lambda \in [\lambda_-, \lambda_+]$, where $\lambda_-$ and $\lambda_+$ are the minimum and maximum eigenvalues of $R$, respectively, and are given by

$$\lambda_{\pm} = \left(1 \pm \sqrt{1 - \frac{1}{Q^2}}\right)^2.$$  \hspace{1cm} (7)

The distribution of the components $\{u_{kl}^k; l = 1, \ldots, N\}$ of eigenvector $u^k$ of a random correlation matrix $R$ has also been formulated in previous studies. It has been shown that the distribution conforms a Gaussian distribution with zero mean and unit variance,

$$\rho_{rm}(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$  \hspace{1cm} (8)

1.4. Matrix Decomposition

The correlation matrix $C$, as expressed in Equation 3 and 4, can be decomposed to filter out the effects of random noises and isolate the correlations characterized by the deviating eigenvalues. Having calculated the complete set of eigenvalues and eigenvectors, the expansion of the correlation matrix $C$ is of the form

$$C = \sum_{\alpha=1}^{N} \lambda_{\alpha} |\alpha\rangle\langle\alpha|.$$  \hspace{1cm} (9)

2. Method

In this research, we collected the daily closing price of $N = 267$ stocks belonging to the Indonesian Stock Exchange (IDX) from January 1, 2007, to October 28, 2014 ($L \approx 2025$ trading days), which is publicly accessible from the website (http://finance.yahoo.com).

From the dataset, we then construct the correlation matrix and calculate the eigenvalues and eigenvectors. The eigenvalue distribution of the real data will be compared with that of the random correlation matrix as expressed in Equation 6 to distinguish genuine information from noise. The largest eigenvector will be further investigated to infer reasonable interpretation of the largest eigenvalue. Then, the components of the next largest deviating eigenvectors will be listed and examined to identify the correlated groups and the business sectors in which it belongs. From these results, the general structure of the stock market system could be described. Having identified the modes and groups of the system, we then quantify the contribution of these various modes by decomposing the correlation matrix into three parts – market-wide effect, correlated groups, and random noise – to see the underlying dynamics of the stock market.

3. Result and discussion

3.1. Eigenvalue Distribution of the Correlation Matrix

The correlation matrix of the IDX stock price data is constructed using Equation 4, and the distribution $P(C_{ij})$ of the elements $\{C_{ij}; i < j\}$ of the correlation matrix is then plotted in Figure 1 and shows a slightly positive mean and a heavy positive tail. Overall, this result shows that constructive behaviour is more dominant than anti-correlated behaviour among the stocks as generally seen in other stock markets; however, a mean value of $\langle C_{ij} \rangle = 0.0527$ is considerably small compared to developed markets throughout the world.
Figure 1. Distribution $P(C_{ij})$ of elements of the correlation matrix $C$ constructed from the daily closing price returns of 267 stocks for the period 2007 – 2014 (blue line), with a mean value of $\langle C_{ij} \rangle = 0.0527$. The symmetric curve shows the distribution of the correlation coefficients for the random matrix or "control" $P(R_{ij})$ (green line), which is consistent with a Gaussian distribution with zero mean.

Next, we find the eigenvalues $\lambda_i$ for $i = 1, \ldots, 267$ of the correlation matrix $C$. Figure 2a shows the distribution of the eigenvalues of $C$ being compared with the distribution $P_{rm}$ of the eigenvalues of the random correlation matrix $R$ as defined by Equation 6. By using Equation 7, we find the lower and upper bounds of $P_{rm}$ which are 0.405 and 1.858, respectively. The result shows that a majority of the eigenvalues of $C$ lies within the theoretical bounds of the random matrix theory. This indicates that the modes corresponding to these small eigenvalues are mere

Figure 2. Eigenvalue distribution $P(\lambda)$ for $C$ constructed from the daily closing price returns of 267 stocks for the period 2007 – 2014 (red line) in comparison with the RMT result $P_{rm}(\lambda)$ (blue curve) of Equation 6. (a) The bulk of the eigenvalue distribution of $C$ lies under the $P_{rm}(\lambda)$ curve. (b) The deviation of the largest eigenvalue is significant where $\lambda_1 > 10\lambda_+$. 
consequence of a random process. Using this method, we could distinguish noise – properties of \( C \) which conform with that of the random matrix – from genuine information – properties of \( C \) which deviate – in the stock data.

The area of interest is the deviating eigenvalues which can be seen in both extremes of the eigenvalue distribution. However, for this specific research, only the large eigenvalues will be further examined. As seen in Figure 2b, We find that the largest eigenvalue \( \lambda_1 \) is 24.578, which is over ten times larger than the theoretical upper bound of the random matrix theory. The next few largest eigenvalues are 3.416, 2.743, 2.377, and 2.311.

![Distribution of the eigenvector components](image)

**Figure 3.** Distribution \( \rho(u) \) of the components of (a) the largest eigenvector, (b) third largest eigenvector, and (c) eigenvectors within the bulk \( \lambda_- < \lambda < \lambda_+ \), compared with the Gaussian distribution from the RMT prediction of Equation 8 (red line).

### 3.2. Statistics of the Eigenvector

To further confirm the uniqueness of the deviating eigenvalues and the randomness of the remaining eigenvalues, we look at their eigenvector components. For eigenvalues within the theoretical bound, the distribution of the components of the corresponding eigenvector forms a Gaussian distribution (Figure 3c). This result shows that these eigenvectors are similar to the eigenvectors of a random correlation matrix whose distribution is expressed in Equation 8. Whereas for eigenvalues above the theoretical bound, the distribution of the components of the corresponding eigenvectors shows heaviness to either the positive or negative tail (Figure 3b). Notably, the largest eigenvector shows a characteristic trend which is also present in other
Figure 4. Pixel representation of the overlap matrix where black pixel indicates an inner product of one and white pixel indicates an inner product of zero.

stock markets around the world. All the components of the largest eigenvector have the same sign (Figure 3a), indicating a common and collective movement of all stocks. The next largest eigenvectors show localization, meaning that only a few stocks contribute in these eigenvectors. This result shows that these deviating eigenvectors give information about the correlated groups within the stock market.

We also examine the stability of the deviating eigenvectors throughout time by using the ”overlap matrix”. Firstly, we divide the original data into two parts based on their chronological order and calculate the correlation matrix $C$ for both sets of data. Next, we identify the $p$ eigenvectors corresponding to the $p$ largest eigenvalues which deviate from the RMT upper bound $\lambda_+$. We then construct a $p \times N$ matrix $D$ with elements $D_{kj} = \{u^k_j; k = 1, ..., p; j = 1, ..., N\}$. Finally, we compute a $p \times p$ overlap matrix $O(t, \tau) = DA^T_B$, where elements $O_{ij}$ is the scalar or inner product of eigenvector $u^i$ from the first half of the data with $u^j$ from the second half. If all the $p$ deviating eigenvectors are stable in time, $O_{ij} = \delta_{ij}$.

3.3. Interpretation of the Largest Eigenvalue and Its Corresponding Eigenvectors

Since all components of the largest eigenvector have the same sign as seen in Figure 3a, the largest eigenvalue represents an influence that is common to all stocks, and thus, the largest eigenvector quantitatively shows that certain newsbreaks affect all stocks. In real situations, influences that affect all stocks alike are usually national events, such as economic growth and interest rate increase, or world events, such as wars and civil unrest. These types of events may inevitably have an impact on the buying power and demand of investors, and consequently, the prices of stocks are also affected. In summary, all of these effects which influence the market as a whole will be termed ”market-wide effect”, and the largest eigenvalue and eigenvector will be referred to as the market mode.

To strengthen our previous finding concerning the largest eigenvalue as the market mode, we compare the projection (scalar product) of the time series $S_i(t)$ on the eigenvector $u^1$ with the standard measure of Indonesian stock market performance – the JKSE index. We calculate the projection $S^{(1)}(t)$ of the time series $S_i(t)$ on the eigenvector $u^1$,

$$S^{(1)}(t) \equiv \sum_{i=1}^{267} u^1_i S_i(t). \quad (10)$$

By definition, $S^{(1)}(t)$ quantifies the scaled price index defined by $u^1$. Figure 5b shows $S^{(1)}(t)$ regressed against $S_{JKSE}$, which appears as a narrow scatter around a linear fit with correlation
**3.4. Interpretation of Deviating Eigenvectors** $\mathbf{u}^2 - \mathbf{u}^5$

As we have seen before in Figure 3b, the deviating eigenvectors $\mathbf{u}^2 - \mathbf{u}^5$ are localized; thus, we can calculate the number of significant contributors within the eigenvector. Furthermore, we can identify the significant contributors to yield an interpretation regarding the economic relevance of these eigenvectors.

Before we sort the components of each deviating eigenvector and interpret its meaning, we note that the largest eigenvalue is an order of magnitude larger than the others, and thus, it tends to constrain the remaining $N-1$ eigenvalues. Therefore, in order to examine the deviating eigenvectors, we must first remove the effect of the largest eigenvalue $\lambda_1$ or the market mode. We do this by implementing the linear regression

$$G_i(t) = \alpha_i + \beta_i G^{(1)}(t) + \epsilon_i(t),$$

where $G^{(1)}(t)$ is defined similarly with Equation 10 which is $G^{(1)}(t) \equiv \sum_{j=1}^{267} u_j^1 G_{j}(t)$. We then recompute the correlation matrix $\mathbf{C}$ using the residuals $\epsilon_i(t)$ with Equation 2 and 3; and next, we compute the eigenvectors $\mathbf{u}^k$ of $\mathbf{C}$ thus obtained for their significant participants to be examined.

In the second largest eigenvector $\mathbf{u}^2$, we find that the ten largest components are mainly stocks from agricultural and mining sector. However, the rest of the deviating eigenvectors do not belong to a specific sector as can be seen in Table 1. Figure 6 also shows the components of the stocks within the deviating eigenvectors, and it could be seen that eigenvectors $\mathbf{u}^3$ to $\mathbf{u}^5$ is inconclusive and not well-defined. Nevertheless, we could still conclude that there are two main correlated groups that manifest in the Indonesian stock market which are the natural resource companies represented by $\mathbf{u}^2$ and the non-natural resource companies which deal with finance, services, and goods represented by the eigenvectors $\mathbf{u}^3$ to $\mathbf{u}^5$. This result seems acceptable, remembering that Indonesia is one of the leading producers of natural resource, and thus, companies from this sector have been the vanguard of the economic development of Indonesia as can be seen from the dominance of such companies in the second largest eigenvectors.

---

**Figure 5.** (a) Projection of $S_i(t)$ on the largest eigenvector (blue line) compared with the JKSE index (red line) as benchmark. Both time series manifest the sub-prime mortgage crisis which occurred between 2008 and 2009, showing the vulnerability of IDX to the economic crisis happening worldwide. (a) Linear regression of $S^{(1)}(t)$ against $S_{JKSE}^{(1)}(t)$ yields a slope of 0.994, which is the corresponding correlation coefficient value.
Table 1. Ten largest components of the eigenvectors $u^2$ to $u^5$. The columns show ticker symbols, company name, and industry sector, respectively.

| Ticker | Company Name                  | Sector                              |
|--------|-------------------------------|-------------------------------------|
| LSIP   | London Sumatra Indonesia      | Agriculture                         |
| UNSP   | Bakrie Sumatera Plantations   | Agriculture                         |
| AALI   | Astra Agro Lestari            | Agriculture                         |
| TBLA   | Tunas Baru Lampung            | Agriculture                         |
| SGRO   | Sampoerna Agro                | Agriculture                         |
| PTBA   | Tambang Batubara Bukit Asam   | Mining                              |
| TINS   | Timah (Persero)               | Mining                              |
| BUMI   | Bumi Resources                | Mining                              |
| ELTY   | Bakrieland Development        | Property, real estate, and construction |
| MEDC   | Medco Energi internasional    | Mining                              |
| ENRG   | Energi Mega Persada           | Mining                              |
| BCIC   | Bank Mutiara                  | Finance                             |
| LMAS   | Limas Indonesia Makmur        | Trade, Service, and Investment      |
| UNSP   | Bakrie Sumatra Plantations    | Agriculture                         |
| BMTR   | Global Mediacom               | Trade, services, and investment     |
| BIPP   | Bhuwanatala Indah Permai     | Property, real estate, and construction |
| MTDL   | Metrodata Electronics         | Trade, services, and investment     |
| BKSL   | Sentul City                   | Property, real estate, and construction |
| TRUB   | Truba Alam Mamunggal Engineering | Infrastructure, utilities, and transportation |
| ELTY   | Bakrieland Development        | Property, real estate, and construction |
| SSIA   | Surya Semesta Internusa       | Property, real estate, and construction |
| PANS   | Panin Sekuritas               | Finance                             |
| BMRI   | Bank Mandiri (Persero)        | Finance                             |
| ALMI   | Alumindo Light Metal Industry | Basic industry and chemicals        |
| BBCA   | Bank Central Asia             | Finance                             |
| CTRA   | Ciputra Development           | Property, real estate, and construction |
| KIJA   | Kawasaki Industri Jababeka    | Property, real estate, and construction |
| UNSP   | Bakrie Sumatra Plantations    | Agriculture                         |
| KAEF   | Kimia Farma (Persero)         | Consumer goods industry             |
| WEHA   | Panorama Transportasi         | Infrastructure, utilities, and transportation |
| INKP   | Indah Kiat Pulp and Paper     | Basic industry and chemicals        |
| TKIM   | Pabrik Kertas Tjiwi Kimia     | Basic industry and chemicals        |
| CPRO   | Central Proteina Prima        | Agriculture                         |
| ICBP   | Indofood CBP Sukses Makmur    | Consumer goods industry             |
| INDF   | Indofood Sukses Makmur        | Consumer goods industry             |
| SPMA   | Suparma                       | Basic industry and chemicals        |
| MNCN   | Media Nusantara Citra         | Trade, services, and investment     |
| BMTR   | Global Mediacom               | Trade, services, and investment     |
| AUTO   | Astra Otoparts                | Miscellaneous industry              |
| INTA   | Intraco Penta                 | Trade, services, and investment     |
Figure 6. The normalized eigenvector components $u_i$ of stock $i$ corresponding to the second to fifth largest eigenvalues $\lambda_2 - \lambda_5$ of the correlation matrix. The stocks are sorted by industrial sectors, 1: agriculture; 2: basic industry and chemicals; 3: consumer goods industry; 4: finance; 5: infrastructure, utilities, and transportation; 6: mining; 7: miscellaneous industry; 8: property, real estate, and building construction; 9: trade, services, and investment, which are separated by dashed lines.

The deviating eigenvectors we have discussed represent the correlated groups and are defined as the "normal modes” of the stock market movement. In a physical system, normal mode is usually associated with a pattern of motion in which all parts of the system move with the same frequency. All other forms of motion of the system can be described as the linear combination of the normal modes. Analogously, the normal modes of the stock market, which are the correlated groups, describe the most basic or simplest form of movement. All movements or motions of the system are linear combinations of the normal modes.

3.5. Decomposition of the Correlation Matrix

To answer the lack of formation of sector classes from the eigenvectors, we decompose the correlation matrix into the following: the market-wide, group, and random part. The matrix decomposition is carried out by implementing Equation 9 adjusted to satisfy our objective.

$$C = C_m + C_g + C_r$$  (12)

$$C = \lambda_1 |1\rangle\langle 1| + \sum_{\alpha=2}^{5} \lambda_\alpha |\alpha\rangle\langle \alpha| + \sum_{\alpha=6}^{267} \lambda_\alpha |\alpha\rangle\langle \alpha|.$$  (13)

Having obtained $C_m$, $C_g$, and $C_r$, we then plot the distribution of the elements of the three correlation matrices. Figure 7 shows the correlation matrix element distribution of $C_m$, $C_g$, and $C_r$ for IDX and also S&P 500 as comparison. From this result, we see the difference between the characteristics of a developed market from the U.S. and an emerging market, such as IDX.
In the U.S. market, the group part which is constructed from the deviating eigenvectors and therefore responsible for the formation of the correlated groups has a fat tail on the positive side indicating that a large portion of the positive correlation is from the correlating clusters or sectors. Whereas in the IDX case, the group part is still very similar to the random part, with only a small positive tail. This indicates that IDX is still highly driven by only the market-wide effect. It has also been showed previously in the overlap matrix before that only the first eigenvector is stable throughout time.

4. Conclusion
In conclusion, The IDX stock price data show high agreement with the random matrix with more than 90% of the eigenvalues lying inside the theoretical bound. Thus, we are able to distinguish noise from genuine information. We have also successfully identified the normal modes of the stock market system which are the vector basis of all the possible movements of the market. Finally, there are indications of the formation of two clusters, which are the natural resource and non-natural resource industries. However, decomposition of the correlation matrix shows that the formation of the correlated clusters is still premature, and the correlations that exist among the stocks are still dominated by the market-wide effects. These are two characteristics which are common to various emerging markets throughout the world, and therefore, the Indonesian Stock Exchange, IDX, is also classified as an emerging market.

References
[1] Mantegna R N and Stanley H E 2000 An Introduction to Econophysics: Correlations and Complexity in Finance (Cambridge: Cambridge University Press)
[2] Plerou V, Gopikrishnan P, Rosenow B, Nunes Amaral L A N, Guhr T and Stanley H E 2002 Phys. Rev. E 65 066126
[3] Laloux L, Cizeau P, Bouchaud J P and Potters M 1999 Phys. Rev. Lett. 83 1467
[4] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N and Stanley H E 1999 Phys. Rev. Lett. 83 1471
[5] Gopikrishnan P, Rosenow B, Plerou V and Stanley H E 2001 Phys. Rev. E 64 034106(R)
[6] Kim D H and Jeong H 2005 Phys. Rev. E 72 046133
[7] Pan K and Sinha S 2007 Phys. Rev. E 76 046116
[8] finance.yahoo.com