

Research Article

RMT Assessments of the Market Latent Information Embedded in the Stocks' Raw, Normalized, and Partial Correlations

Dror Y. Kenett, Yoash Shapira, and Eshel Ben-Jacob

*School of Physics and Astronomy, The Raymond and Beverly Sackler Faculty of Exact Sciences,
Tel-Aviv University, Tel-Aviv 69978, Israel*

Correspondence should be addressed to Eshel Ben-Jacob, eshelbj@gmail.com

Received 3 September 2009; Accepted 10 December 2009

Recommended by A. Thavaneswaran

We present here assessment of the latent market information embedded in the raw, affinity (normalized), and partial correlations. We compared the Zipf plot, spectrum, and distribution of the eigenvalues for each matrix with the results of the corresponding random matrix. The analysis was performed on stocks belonging to the New York and Tel Aviv Stock Exchange, for the time period of January 2000 to March 2009. Our results show that in comparison to the raw correlations, the affinity matrices highlight the dominant factors of the system, and the partial correlation matrices contain more information. We propose that significant stock market information, which cannot be captured by the raw correlations, is embedded in the affinity and partial correlations. Our results further demonstrate the differences between NY and TA markets.

Copyright © 2009 Dror Y. Kenett et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Stock markets behave as complex dynamic systems, and as such, it is critical to investigate the dependencies (interactions) between the dynamics of the system variables (stocks, bonds, etc.). It is common to associate such interactions with the notion of correlation, or similarity. Indeed, much effort is dedicated to study and understand such stock cross-correlations [1, 2] in an attempt to extract maximum market latent information that is embedded in the interactions between the market variables.

Extraction of relevant information from the empirical cross-correlations between stocks is difficult, as has been noted by Plerou et al. [3, 4]. To discern to what extent the empirical stock correlation coefficients are a result of random noise rather than important system information, they proposed to make use of Random Matrix Theory (RMT). RMT is a tool that was originally developed and used in nuclear physics by Wigner and Dirac [5], Dyson [6], and Mehta [7] in order to explain the statistics of the energy levels of complex quantum systems. Recently, it has also been applied to studies of economic data [3, 4, 8–20].

Whereas the main focus of RMT analysis of stock data has been on the analysis of stock cross-correlation matrices, here we assess the importance of RMT investigations for two alternative similarity measures. The first is a measure of normalized correlations, computed by a collective normalization of the correlation matrix, which has been termed the affinity transformation [2, 21, 22]. The second is a measure of partial correlations [23], which provides a way to study the residual stock correlations, after removing the mediating effect of the index. This measure was chosen following our previous work [2], where we have used this measure to show the influence of the Index on stock correlations.

Here we analyze financial time series from the New York stock market (S&P500, available at [24]) and the Tel-Aviv stock market (TASE, available at [25]). For the S&P stock data, we studied a set of 445 stocks, for the trading time period of 03/01/2000 to 15/003/2009. In addition, we also investigated the Dow Jones Industrial Average (DJIA) index and its constituting stocks [24], which are also members of the S&P500 index. For the TA data, we only used stocks with substantial capital and substantial daily turnover for the same time period. For each stock, we use the daily adjusted closing price to compute the stock return, using the standard transformation

$$r_i(t) = \ln \left[\frac{Y_i(t + \Delta t)}{Y_i(t)} \right] = \ln[Y_i(t + \Delta t)] - \ln[Y_i(t)], \quad (1.1)$$

where $Y_i(t)$ is the daily adjusted closing price of stock i , and the time interval Δt was taken to be equal to 1 trading day. The result of the screening process [2] is 445 stocks and the S&P500 index for the S&P dataset, 30 stocks and the DJIA index for the DJIA dataset, and 26 stocks and the TA25 index, for the TA dataset.

The comparison between the different similarity measures is performed employing the commonly used RMT. The distribution of the eigenvalues of each matrix is evaluated and compared to the expected distribution of eigenvalues of the corresponding random matrix (the original matrix after random shuffling of the elements, also known as the Wishart matrix [3]).

2. Stock Correlation Metrics

The three similarity measures we use in this work are: (1) Raw correlations $C(i, j)$ —the Pearson's pairwise correlation between stocks i and j ; (2) Affinity correlations $\text{Aff}(i, j)$ —the normalized correlation between stocks i and j , according to the correlation of each one with all other stocks, computed by the affinity collective normalization [2, 21, 22]; (3) Partial correlations $\rho(ij, m)$ —the residual correlation between stocks i and j , after removing the mediating effect of a stock m .

2.1. Raw Correlation Matrices

The stock raw correlations are calculated using the Pearson correlation coefficient

$$C(i, j) = \frac{\langle r(i) - \mu(i) \rangle \langle r(j) - \mu(j) \rangle}{\sigma(i)\sigma(j)}, \quad (2.1)$$

where $r(i)$ and $r(j)$ are the return of stock i and j , $\mu(i)$ and $\mu(j)$ denote the corresponding means, $\sigma(i)$ and $\sigma(j)$ are the corresponding standard deviations (STDs), and $\langle \rangle$ denotes the average over time. Note that $C(i, j)$ is a symmetric square matrix and $C(i, i) = 1$ for all i .

2.2. Normalized Correlation (Affinity) Matrices

The matrices are normalized using the affinity transformation, a special collective normalization procedure first proposed by Baruchi et al. [21, 22]. The idea is to normalize the correlations between each pair of stocks according to the correlations of each of the two stocks with all other stocks. This process is in fact a calculation of the correlation of correlations or metacorrelation. The metacorrelations $MC(i, j)$ are the Pearson's correlation between rows i and j in the correlation matrix after reordering. In the reordering process, the elements $C(i, i)$ and $C(j, j)$ are taken out. The correlation vector for i is $\{C(i, j), C(i, 1), C(i, 2), \dots\}$ and for j it is $\{C(i, j), C(j, 1), C(j, 2), \dots\}$,

$$MC(i, j) = \frac{\sum_{k \neq i, j}^N (C(i, k) - \mu C(i)) * (C(j, k) - \mu C(j))}{[\langle C(i)^2 \rangle * \langle C(j)^2 \rangle]^{1/2}}. \quad (2.2)$$

In other words, the metacorrelation is a measure of the similarity between the correlations of stock i with all other stocks to the correlations of stock j with all other stocks. Using the metacorrelations, the normalized correlations $Aff(i, j)$ are

$$A(i, j) = \sqrt{C(i, j) \cdot MC(i, j)}. \quad (2.3)$$

The affinity transformation process emphasizes subgroups of variables (stocks) in the system, by removing the effect of the background noise of correlation. Groups (clusters) identified in the affinity matrix are significant in the system, and warrant further investigation. We demonstrate the strength of the affinity transformation in Figure 1, where we compare the S&P500 dataset correlation matrix to its affinity matrix. Both matrices are ordered similarly, and the groups weakly visible in the correlation matrix (Figure 1(a)) are emphasized and highlighted by the affinity transformation process (Figure 1(b)).

2.3. Partial Correlation Matrices

Recently [2], the concept of partial correlation has been applied in the study of financial data. A partial correlation is a statistical measure indicating how a third mediating variable affects the similarity between two variables [23]. The partial correlation coefficient is given by

$$\rho(ij, m) = \frac{C(i, j) - C(i, m)C(j, m)}{\sqrt{[1 - C^2(i, m)][1 - C^2(j, m)]}}, \quad (2.4)$$

where $C(i, j)$ is the pairwise correlation between stock i and j , $C(i, m)$ is the pairwise correlation between stock i and the index m , and $C(j, m)$ is the pairwise correlation between stock j and the index.

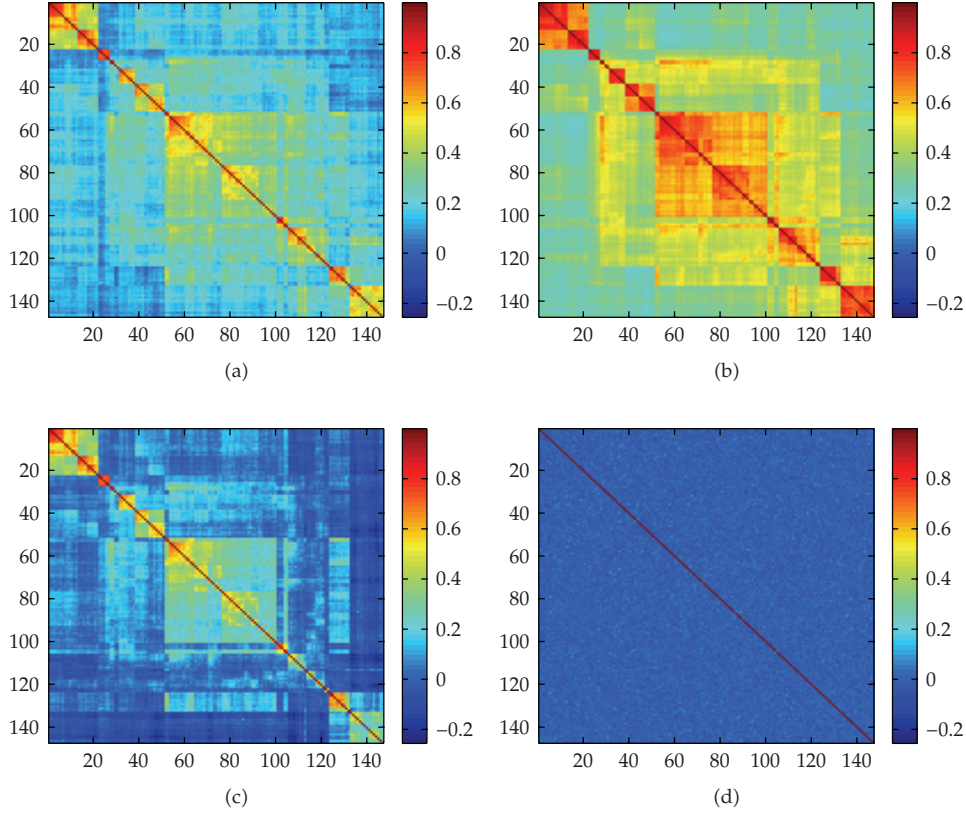


Figure 1: Raw correlation (a), affinity (b), and partial correlation (c) matrices of a representative set of 147 stocks from the S&P dataset. A random matrix (d) is created using size similar properties.

As we have recently reported [2], computing the partial correlations between stocks after removing the mediating effect of the index reveals the naked structure of the market. Thus, here we will repeat this process and compute the stock partial correlation matrices for the three datasets, using the index as the mediating variable.

3. Random Matrix Theory (RMT)

Using RMT, we study whether there are eigenvalues that deviate from the range of the random ones (the eigenvalues of the corresponding random matrix). The eigenvalues of the random matrix represent the limit of zero-embedded information (maximum entropy), hence the deviation from this range provides a measure for relevant latent information about the market organization that is embedded in the correlation matrices.

3.1. Eigenvalues of Random Matrix

Let us consider a portfolio of N stocks, each with T time records. If these time series are uncorrelated, then the resulting random matrix (also known as the Wishart matrix [3]) has

eigenvalues that follow the distribution

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}, \quad (3.1)$$

with $n \rightarrow \infty$, $T \rightarrow \infty$, and $Q = T/n \geq 1$. This distribution is bounded by

$$\lambda_{\max} = \sigma^2 \left(1 + \frac{1}{Q} + 2\sqrt{\frac{1}{Q}} \right), \quad (3.2)$$

$$\lambda_{\min} = \sigma^2 \left(1 + \frac{1}{Q} - 2\sqrt{\frac{1}{Q}} \right). \quad (3.3)$$

The spectral density is different from zero in the interval $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$ and in the case of correlation matrices, $\sigma^2 = 1$ (other values for σ^2 have been proposed in the past, see, e.g., [11, 13]). Using the distribution presented in (3.1), we study the eigenvalue distribution for the different sets of stocks.

3.2. Assessment of the Embedded Information

We compare the eigenvalues statistics and scaling of the raw, affinity and partial correlation matrices, to those of the corresponding random matrices. It has been proposed in the past by Plerou et al. [3, 4], Laloux et al. [14], Garas and Argyrakis [11], and others that eigenvalues deviating from the eigenvalue regime of the random matrix, that is, $\lambda_{\min} < \lambda < \lambda_{\max}$, contain the relevant economic information stored in the correlation matrices.

3.3. Zipf Plot of the Eigenvalues and Power Law Scaling

While inspection of the eigenvalue distribution was found to be useful, it cannot be used for the case of small markets, and it does not always capture important information embedded in the scaling behavior of the leading eigenvalues. To overcome these difficulties, we investigated the scaling of the eigenvalues. Following the notion of a Zipf law that was found in text [26, 27], we simply plot the values of the ordered eigenvalues as function of their rank using a log-log scale. This simple presentation was found to be quite powerful in distinguishing between the different similarity measures and revealed a power law-scaling behavior for the correlation matrices.

3.4. Eigenvalue Spectrum

To further study the scaling behavior of the eigenvalues in this context, we first order the eigenvalues such that $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n$ and study the scaling of "frequency" spectrum of the eigenvalues, defined by

$$\omega(s) = \lambda(s) - \lambda(s-1), \quad s = 1, 2, \dots, n. \quad (3.4)$$

Once we compute the frequency spectrum of the eigenvalues, ω , we study the distribution of this spectrum.

Table 1: Value of Q , λ_{\min} , and λ_{\max} for the random matrices created for the S&P, DJIA, and TA datasets.

	Q	λ_{\min}	λ_{\max}
S&P	4.9438	0.3028	2.1018
DJIA	77.0333	0.7851	1.2409
TA	86.9231	0.7970	1.2260

4. Results

4.1. Stock Similarity Matrices

We begin by calculating the correlation, affinity, and partial correlation matrices for each of the three datasets. Then, to create the random matrices, we shuffled the stock return data, for each of the three datasets separately. We repeated this shuffling 10 times and averaged the results into one single random matrix. The Q value and maximal and minimal eigenvalue computed for these shuffled matrices are summarized in Table 1.

An example of the different correlation matrices for a representative set of 147 stocks of the S&P data, and the corresponding random matrix, is presented in Figure 1. Throughout this section, we will demonstrate our results using the S&P dataset.

4.2. The Eigenvalue Distributions

For each similarity matrix we compute its eigenvalues and then investigate their distribution. In Figure 2 we plot the probability distribution function (PDF) of the eigenvalues for the three similarity matrices (for the S&P dataset). We compare each PDF to the corresponding distribution of eigenvalues of the random matrix. The results obtained for the raw correlations are similar to those reported by Plerou et al. [3], who analyzed a similar set of stocks.

First, we compare the eigenvalues of the raw and affinity correlations (Figures 2(a) and 2(b)). We find that for both similarity measures, the first eigenvalue (the principal eigenvalue) is significantly larger than the rest (~ 10 times larger for both cases). The principal eigenvalue is commonly associated with the market [3, 4, 11]. Comparing these two sets of eigenvalues, we note that the normalization process used to compute the affinity matrices results in a much larger principal eigenvector ($\lambda = 131, 233$ for the correlation and affinity matrix, resp.). Furthermore, in the case of the affinity matrices, there are less deviating eigenvalues above λ_{\max} , for all three cases (Table 2).

Next, we turn to study the eigenvalues of the partial correlation matrices (Figure 2(c)). Studying this distribution, it is possible to observe that there are more eigenvalues that deviate from the noise, especially ones that are larger than λ_{\max} (Table 2). Another immediate observation is that the principal eigenvalue is now significantly smaller than in the case of the correlation and affinity matrices. However, it is important to keep in mind that in this case the principal eigenvalue is no longer associated with the market, since the effect of the index (which is a representative of the market [2]) has been removed.

For the DJIA and TA datasets, we focus on the number of deviating eigenvalues from λ_{\max} of the corresponding random matrix (for each dataset separately), for the raw correlation, affinity, and partial correlation matrices (Table 2). Here again we observe more deviating eigenvalues in the case of the partial correlations. For the raw correlation matrix, there are 3 and 2 such eigenvalues for the DJIA and TA datasets, respectively, and 2 and 1

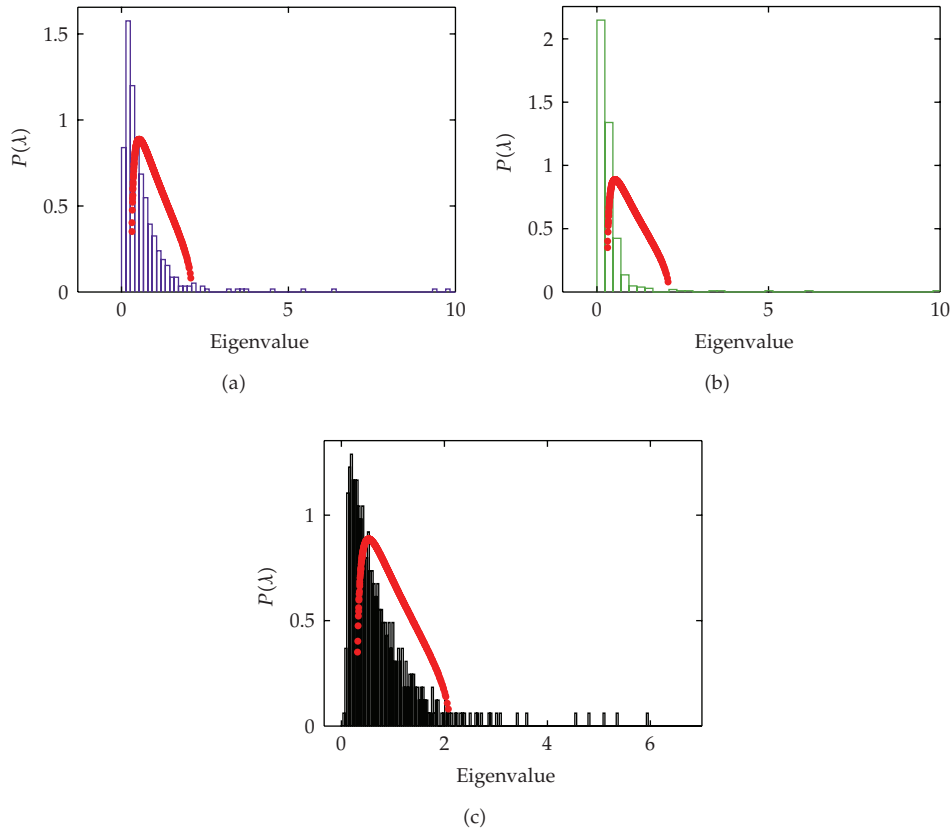


Figure 2: Probability Distribution Function (PDF) of the eigenvalues of the similarity measures for the S&P dataset, for the raw correlation (a), affinity (b), and partial correlation (c). The distributions are presented using straight lines, and the distributions for the corresponding random matrices are presented using red circles. In the case of the raw correlation and affinity matrices, the first eigenvalue (principal eigenvalue (PE)) is significantly larger than the other eigenvalues (131 and 233, resp.). In the case of the partial correlation, we see that there are more eigenvalues that are larger than the largest theoretical eigenvalue λ_{\max} (Table 2), and the principal eigenvalue is now much smaller (33).

Table 2: Number and percentage (out of total) of empirical eigenvalues that deviate from the theoretical maximal eigenvalue, $\lambda > \lambda_{\max}$, for the raw correlation, affinity, and partial correlation matrices for the three datasets.

	Raw correlation	Affinity	Partial correlation
S&P	18 (4.04%)	13 (2.92%)	28 (6.29%)
DJIA	3 (10.00%)	2 (6.67%)	6 (20.00%)
TA	2 (7.69%)	1 (3.85%)	5 (19.23%)

for the affinity matrices. These include the principal eigenvalue. In contrast, in the case of the partial correlations, there are 6 and 5 such eigenvalues for the DJIA and TA datasets, respectively. This is in agreement with the results of the S&P dataset, where we observed more eigenvalues that deviate above λ_{\max} for the case of the partial correlation matrix.

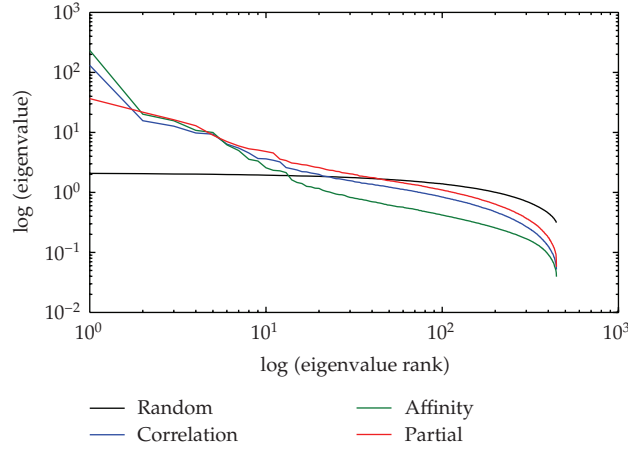


Figure 3: Zipf plot of the S&P dataset. Here we plot the log of the value of the eigenvalue as function of the log of its rank. A power law-like behavior is observed for the raw (blue), affinity (green), and partial (red) correlation, in comparison to the random matrix (black).

4.3. Scaling Behavior of Stock Similarity Measures

To further investigate the eigenvalues of the different similarity matrices, we order the eigenvalues by rank and investigate whether they follow a Zipf's law like behavior [26, 27].

To test if the eigenvalue distributions display a Zipf Law behavior, we calculate the slope of each distribution presented in Figure 3. For the first 100 eigenvalues, we see a power law-scaling behavior, which disappears for smaller eigenvalues. We find that the slope equals -0.9350 , -1.2582 , -0.7999 , and -0.0699 for the raw correlation, affinity, partial correlation, and random matrices, respectively.

4.4. Eigenvalue Spectra

Comparison of the eigenvalues spectra, as defined by ω in (3.3), is presented in Figure 4. For all three-similarity measures, it is clear that the first 10 values of ω are significantly different from those of the corresponding random matrices.

In Figure 5 we plot the ω distribution for each of the similarity measures against the ω distribution for the eigenvalues of the random matrix. Studying both Figures 4 and 5, we observe that as was discussed in the previous sections, for all three similarity measures we find eigenvalues that deviate from noise.

5. Eigenvalue Spectral Entropy

The concept of eigenvalue entropy was used as a measure to quantify the deviation of the eigenvalue distribution from a uniform one [28]. The idea was first used in the context of biological systems [29, 30]. Here we use a similar concept to quantify the ω spectra rather than the eigenvalue distribution. The spectral entropy, SE, is defined as

$$SE \equiv -\frac{1}{\log(N)} \sum_{i=1}^N \Omega(i) \log[\Omega(i)], \quad (5.1)$$

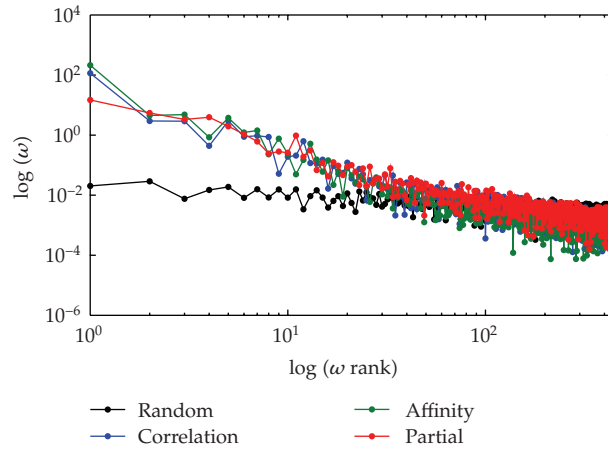


Figure 4: ω spectrum of the raw (blue), affinity (green), and partial (red) correlation, in comparison with the corresponding random matrix (black), for the S&P dataset.

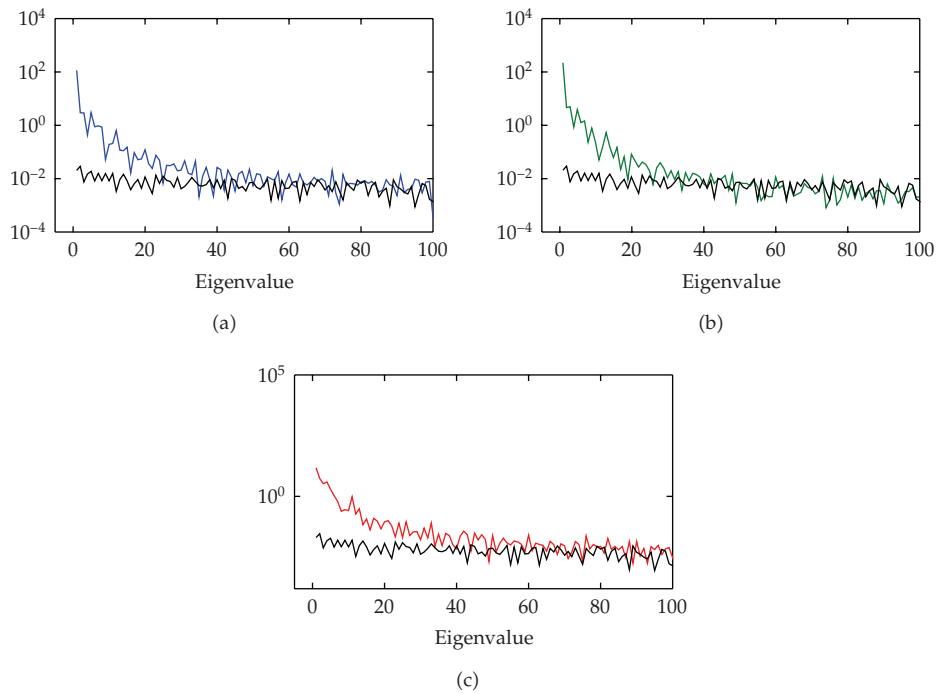


Figure 5: Comparison of the ω spectra for the raw ((a), blue), affinity ((b), green), and partial ((c), red) correlations, to the corresponding random matrix case (black). Data used is of the S&P data, using a semilog plot. Only the first 100 values of ω are presented.

where $\Omega(i)$ is given by

$$\Omega(i) = \frac{\omega(i)^2}{\sum_{i=1}^N \omega(i)^2}. \tag{5.2}$$

Table 3: Spectral Entropy of for the three different datasets, for each of the four similarity cases.

	Raw correlation	Affinity	Partial correlation	Random
S&P	0.0031	0.0020	0.1378	0.8495
DJIA	0.0199	0.0071	0.4336	0.7418
TA	0.0317	0.0041	0.3979	0.7547

Table 4: Spectral Entropy for the three different datasets, for each of the four similarity cases. For the raw correlation and affinity, the first eigenvalue was removed.

	Raw correlation	Affinity	Partial correlation	Random
S&P	0.2496	0.2315	0.1378	0.8495
DJIA	0.5268	0.3545	0.4336	0.7418
TA	0.0321	0.0042	0.3979	0.7547

Note that the $1/\log(N)$ normalization was selected to ensure that $SE = 1$ for the maximum entropy limit of flat spectra (all ω are equal).

The resulting SE values are presented in Table 3. The lowest entropy reduction (in comparison to that of the corresponding random matrices), hence the highest value of embedded information, was obtained for the affinity matrices.

However, since the first eigenvalue for the raw and affinity correlations includes the effect of the index (the market), we perform additional comparison after first removing the first eigenvalue for the correlation and affinity matrices. The results are presented in Table 4.

Looking at Table 4, we note that in the case of the S&P, the entropy of the partial correlation matrix is now the lowest. This is consistent with the results presented above, derived from the RMT analysis. However, the DJIA and TA datasets display different behavior of the entropy after the removal of first eigenvalue. Comparison of the entropy values for DJIA and TA with and without the first eigenvalue provides a quantification of the index effect on these different types of markets.

6. Discussion

Understanding the similarity between stocks for a given portfolio is of the highest importance. Here we present a comparison of three different similarity measures: raw correlations, affinity (collective normalization) correlations, and partial correlations. The eigenvalue statistics and scaling of each of the matrices are compared with that of the eigenvalues of the corresponding random matrices. The investigation presented here was performed on datasets representing two different markets—the mature and large NYSE, and the young and small TASE.

The affinity matrix, which is the result of a collective normalization of the correlation matrix, has been found to highlight and emphasize cliques and important variables in the system [2, 21, 22]. The most significant difference between the affinity matrices and the raw correlation matrices is that in the affinity case, the first eigenvalue is $\sim 33\%$ larger. Thus, the affinity transformation highlights the impact of the market on the stock correlations.

The partial correlation matrices represent the residual correlations after removing the mediating effect of the index [2]. Garas and Argyrakis [11] used RMT to study the eigenvalue distribution of the Athens Stock Exchange and subtracted from it the contribution of the first

eigenvalue, by normalizing the eigenvalues of the raw correlations. Using partial correlations, as was suggested here, allows to first remove the effect of the index on all the correlations between stocks, and then study the eigenvalue distribution of the matrix.

We found that, in the case of the partial correlations, there are more eigenvalues that deviate from those of the corresponding random matrix (Table 2). It is common to associate the eigenvectors of the deviating eigenvalues with economic sectors (see, e.g., [9]). This has only been partially successful in the past. In the case of the partial correlation matrices, the first eigenvalue is no longer associated with the market mode, since the effect of the market was removed (by eliminating the effect of the index). This led us to the simple intuitive hypothesis that the first eigenvector of the partial correlation matrix will be very similar to the second eigenvector of the raw and normalized correlations. However, this turned out not to be the case, and such similarity was not observed for all three datasets. In the S&P dataset, we found that the first eigenvector of the partial correlation matrix mainly includes stocks from the technology sector, however not exclusively. These findings led us to the conclusion that the association of the eigenvectors of the stock similarity matrices, especially in the case of partial correlations, demands a more rigorous investigation, which will be presented in the future.

To extend our analysis of the eigenvalues of the different similarity measures, we turned to study the full set of eigenvalues of each matrix. Following the notion of Zipf law for text [26, 27], we plotted the values of the ordered eigenvalues as function of their rank, using a log-log scale. This simple presentation was found to be quite powerful in distinguishing between the different similarity measures. The eigenvalue Zipf plot revealed a power law-scaling behavior with a higher slope for the raw and affinity correlations. This analysis tool was found to be especially useful in the investigation of the DJIA and TA datasets.

Next, we investigated the information content in the eigenvalue spectra (in analogy to the energy level gaps, which were originally addressed by random matrix theory). For such a purpose, we used a Shannon entropy-like measure [29, 30] that quantifies the deviations from the limit of maximum entropy (lowest information), which is found for the case of flat spectra. The spectral entropy (SE) revealed that the affinity matrices contain the largest amount of information (minimum spectral entropy). However, with the entropy, after removing the first eigenvalue (that corresponds to the market mode) in the raw correlation, and affinity matrices, we found that the partial correlations contain more information for the S&P stocks. For the DJIA and TA, the affinity correlations contain the largest amount of information, even after the removal of the first eigenvalue. This phenomenon might indicate a weaker effect of the index in the smaller markets; however more work is required to clarify this issue. Furthermore, after removing the first eigenvalue, the entropy of the TA stocks is barely changed, unlike that of the DJIA and S&P stocks. This shows that the two markets (NYSE and TASE) have a different dependency on the market mode. However, to fully understand this observation, it is important to study this phenomenon for other markets, of different sizes and maturity stages.

We have used different measures to investigate the information embedded in different stock similarity matrices. On the one hand, we study the eigenvalues of the different similarity matrices, by comparing them to the regime of random noise, and by studying their scaling behavior and the scaling behavior of their frequencies. On the other, we investigate the information content using a measure of entropy. In the case of the former, we have found that more information is embedded in the partial correlation matrices, where we have found more eigenvalues that deviate from the regime of the noise. In the case of the latter, we have found that more information is embedded in the affinity matrices.

At this stage it is unclear why the two approaches did not fully agree on the most informative similarity measure. A possible explanation is that RMT analysis focuses on a small subset of the eigenvalues, while the entropy analysis focuses on the full set of eigenvalues, and thus the two forms of information are slightly different. However, while each approach emphasized a different similarity measure as being the most informative, neither found the raw correlations to be as such. Thus, we propose the use of the other two similarity measures for the study of stock similarities in a given portfolio.

In conclusion, in this preliminary work we show that important latent market information is embedded in the affinity and partial correlation matrices. This finding has important implications for portfolio optimization and stock risk assessment, and future work should be devoted in these directions. Finally, using these tools to extract meaningful information from stock similarity matrices, we have shown a significant difference between the two types of markets studied here. To fully utilize this result, this work must be expanded to include many other markets. Doing so should elucidate the full picture regarding the difference in information content for different types of markets.

Acknowledgments

The authors would like to thank Rosario Mantegna and Michele Tumminello (Palermo University, Italy) for many fruitful conversations and insights on the subject. They have also greatly benefited from discussions with H. Eugene Stanley (Boston University). Finally, they would like to thank Gitit Gur-Gershgoren (Israel Securities Authority) for many fruitful conversations regarding the Tel Aviv Market. This research has been supported in part by the Tauber Family Foundation and the Maguy-Glass Chair in the Physics of Complex Systems at Tel Aviv University.

References

- [1] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlation and Complexity in Finance*, Cambridge University Press, Cambridge, UK, 2000.
- [2] Y. Shapira, D. Y. Kenett, and E. Ben-Jacob, "The index cohesive effect on stock market correlations," *European Physical Journal B*, vol. 72, no. 4, pp. 657–669, 2009.
- [3] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, "Random matrix approach to cross correlations in financial data," *Physical Review E*, vol. 65, no. 6, Article ID 066126, 18 pages, 2002.
- [4] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, "Universal and nonuniversal properties of cross correlations in financial time series," *Physical Review Letters*, vol. 83, no. 7, pp. 1471–1474, 1999.
- [5] E. P. Wigner and P. A. M. Dirac, "On the statistical distribution of the widths and spacings of nuclear resonance level," *Proceedings of the Cambridge Philosophical Society*, vol. 47, no. 4, p. 790, 1951.
- [6] F. J. Dyson, "The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics," *Journal of Mathematical Physics*, vol. 3, no. 6, pp. 1199–1215, 1962.
- [7] M. L. Mehta, *Random Matrices*, Academic Press, Boston, Mass, USA, 3rd edition, 2002.
- [8] C. Coronello, M. Tumminello, F. Lillo, S. Miccichè, and R. N. Mantegna, "Economic sector identification in a set of stocks traded at the New York Stock Exchange: a comparative analysis," in *Noise and Stochastics in Complex Systems and Finance*, vol. 6601 of *Proceedings of SPIE*, pp. U198–U209, Florence, Italy, May 2007.
- [9] C. Coronello, M. Tumminello, F. Lillo, S. Miccichè, and R. N. Mantegna, "Sector identification in a set of stock return time series traded at the London Stock Exchange," *Acta Physica Polonica B*, vol. 36, no. 9, pp. 2653–2679, 2005.

- [10] J. Daly, M. Crane, and H. J. Ruskin, "Random matrix theory filters in portfolio optimisation: a stability and risk assessment," *Physica A*, vol. 387, no. 16-17, pp. 4248–4260, 2008.
- [11] A. Garas and P. Argyrakis, "Correlation study of the Athens Stock Exchange," *Physica A*, vol. 380, no. 1-2, pp. 399–410, 2007.
- [12] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley, "Quantifying and interpreting collective behavior in financial markets," *Physical Review E*, vol. 64, no. 3, Article ID 035106, 4 pages, 2001.
- [13] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, "Noise dressing of financial correlation matrices," *Physical Review Letters*, vol. 83, no. 7, pp. 1467–1470, 1999.
- [14] L. Laloux, P. Cizeau, M. Potters, and J. P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance*, vol. 3, no. 3, pp. 391–397, 2000.
- [15] K. G. D. R. Nilantha, Ranasinghe, and P. K. C. Malmini, "Eigenvalue density of cross-correlations in Sri Lankan financial market," *Physica A*, vol. 378, no. 2, pp. 345–356, 2007.
- [16] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, "A random matrix theory approach to financial cross-correlations," *Physica A*, vol. 287, no. 3-4, pp. 374–382, 2000.
- [17] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, "Collective behavior of stock price movements—a random matrix theory approach," *Physica A*, vol. 299, no. 1-2, pp. 175–180, 2001.
- [18] M. Tumminello, F. Lillo, and R. N. Mantegna, "Spectral properties of correlation matrices for some hierarchically nested factor models," in *Proceedings of International Conference on Complexity, Metastability and Nonextensivity*, vol. 965 of *AIP Conference Proceedings*, pp. 300–307, Catania, Italy, July 2007.
- [19] A. Utsugi, K. Ino, and M. Oshikawa, "Random matrix theory analysis of cross correlations in financial markets," *Physical Review E*, vol. 70, no. 2, Article ID 026110, 11 pages, 2004.
- [20] D. Wilcox and T. Gebbie, "An analysis of cross-correlations in an emerging market," *Physica A*, vol. 375, no. 2, pp. 584–598, 2007.
- [21] I. Baruchi, D. Grossman, V. Volman, et al., "Functional holography analysis: simplifying the complexity of dynamical networks," *Chaos*, vol. 16, no. 1, Article ID 015112, 16 pages, 2006.
- [22] I. Baruchi, V. L. Towle, and E. Ben-Jacob, "Functional holography of complex networks activity—from cultures to the human brain," *Complexity*, vol. 10, no. 3, pp. 38–51, 2005.
- [23] K. Baba, R. Shibata, and M. Sibuya, "Partial correlation and conditional correlation as measures of conditional independence," *Australian & New Zealand Journal of Statistics*, vol. 46, no. 4, pp. 657–664, 2004.
- [24] <http://finance.yahoo.com/>.
- [25] <http://www.tase.co.il/>.
- [26] G. K. Zipf, *Selected Studies of the Principle of Relative Frequency in Language*, Harvard University Press, Cambridge, Mass, USA, 1932.
- [27] G. K. Zipf, "Semantic frequency list," *American Speech*, vol. 16, no. 1, pp. 43–45, 1941.
- [28] O. Alter, P. O. Brown, and D. Botstein, "Singular value decomposition for genome-wide expression data processing and modeling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 18, pp. 10101–10106, 2000.
- [29] R. Varshavsky, A. Gottlieb, D. Horn, and M. Linial, "Unsupervised feature selection under perturbations: meeting the challenges of biological data," *Bioinformatics*, vol. 23, no. 24, pp. 3343–3349, 2007.
- [30] R. Varshavsky, A. Gottlieb, M. Linial, and D. Horn, "Novel unsupervised feature filtering of biological data," *Bioinformatics*, vol. 22, no. 14, pp. e507–e513, 2006.