EMERGENT GROUP PROPERTIES IN FINANCIAL MARKETS

O.A. Doria A.

Phd. supervisor: Josep Perelló

Departament de Física Aplicada, Universitat Politècnica de Catalunya. Department de Física, Universitat de Barcelona. Barcelona, Spain.

Abstract

In the present we have analyzed the data from 480 companies of the S&P500 using the Random Matrix Theory and the Inverse Participation Ratio, we analyzed the eigenvalues and its respective eigenvectors in order to find structural behaviour in market. We take off the non-random part from the correlation matrix and use the new correlation matrix to calculate the risk for a given portfolio, showing good results as were expected. Also, we give a complex network perspective of the companies, building a network with the Spanning Tree Algorithm, in order to obtain some useful different measure of group behaviour.

1 Introduction

The financial data has been well studied, several market models have been develop since Bachelier (1900) by first time, explained the prices changes as a Brownian Random Motion (BRM) model, based in this, and supplementing some deficiencies of the BRM model, Osborne (1959), Osborne & Murphy Jr. (1984), defined the return difference. Some properties has been discover, as the stylized facts, Chakraborti *et al.* (2009). The majority of this properties has been studied as individual assets, the point of this work is study group properties of financial data, in order to measure relative quantities in group behaviour, not visible in each independent asset.

To start in our aim, is necessary choose a quantity that relate all assets to study, this is the Empirical Correlation Matrix (ECM). The ECM is defined with all correlation between assets, where each element in the matrix is one single correlation, by this, is a symmetric matrix with the each element of diagonal equal to one: self asset correlation.

We will see in section 2.2 and 2.3, that the ECM contains random information, which we can identify using Random Matrix Theory (RMT). The RMT was developed by physicists to predict the energy levels of heavy nuclei ([Wig1, Wig3, Po,BFFMPW]), Miller & Takloo-Bighash (2007), in order to predict general properties of the systems. Since there, many applications has been found for this theory, one of these, is finance.

The ECM will be modeled as a RMT, in order to find which part is random and which is not. In section 2.3 we will compare theoretical results, simulated and market data, in order to determine differences and similarities, to determine nonrandom part of our ECM.

In section 2.4, we define the Probability Density Function of the correlation matrices, ECM and simulated, to determine the difference between both and see that the ECM has a tendency.

Then, we will use the concept of the Inverse Participation Ratio in section 2.5, to determine the amount of randomness of each eigenstate, taking from this, important information about how the companies are grouped and its relevance into the emergent group, as results, our first notion of market cluster.

After the past information, we are ready to discuss the meaning of the eigenstates, section 2.6, taking out the effect of the largest eigenvalues with its respective eigenvector, to study the difference with the unchanged ECM and the simulated random matrix.

To finish the second section, we will use the Markowitz Portfolio Theory, in order to show the applications of RMT, taking off the randomness of the ECM to determine the risk in the portfolio optimization.

To end this work, we have built a network of the market, using the ECM and applying the Minimum Spanning Tree (MST) algorithm to simplify the ECM. The vertexes are the companies and the weight is defined up to the correlation coefficient. Here, we will determine the importance of some assets respect with others, by counting the number of connections after

the simplification. Other group property will result finding different clusters respect of the spectral analysis.

2 Modeling Financial Data

2.1 Preliminaries

To start modeling financial data, first we have to define some quantities:

$$r_t = \log(p_{t+1}) - \log(p_t) \tag{1}$$

with r_t the return and p_t the price at time t. This quantity has a better time scale behaviour than price difference and is used to build financial models in actuality.

The stochastic modern approach about the dynamics of return is give by:

$$dr = \mu dt + \sigma dW \tag{2}$$

Where dW is a Wiener process, with dW follow a normal distribution that $\langle dW \rangle = 0$ and $\langle dW^2 \rangle = 1$, σ is the volatility and μ the mean return.

The ECM is defined from the return of each asset as $\mathbf{C} = HH^\dagger$ where H in this case is the matrix composed by N returns of M different assets, by this, $\mathbf{C}$ has $M(M-1)/2$ independent real numbers entries.

Using the RMT methods, we have to find the eigenvalues and its respectives eigenstates, for a complete develop see Edelman & Rao (2005).

In the limit of N,$M \to \infty$ with fixed ratio $Q = N/M \geq 1$, the density of the eigenvalues is give by:

$$\rho(\lambda_c) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda_c)(\lambda_c - \lambda_{min})}}{\lambda_c} \tag{3}$$

$$\lambda_{min}^{max} = \sigma^2 \left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right) \tag{4}$$

with $\lambda_c \in [\lambda_{min}, \lambda_{max}]$, σ^2 is the variance of elements of HLaloux *et al.* (1999), therefore σ^2 is the average eigenvalue of $\mathbf{C}$. This comes with important implications: the eigenvalues are bounded and greater than zero for $Q \geq 1$ and have a maximum near the lower eigenvalue (λ_{min}).

2.2 Data

To apply the RMT to financial data, we first have to define the ECM as:

$$\mathbf{C}_{ij} = \frac{1}{N} \sum_{k=1}^{N} dX_i(k) dX_j(k) \tag{5}$$

Where the $dX_i = 1/\sigma(dr - \mu dt)$, k the time. The average of return has been subtracted and rescaled to have a constant unit volatility as in Laloux *et al.* (1999).

$\mathbf{C}$ is a symmetric matrix defined as: $-1 < \mathbf{C}_{ij} < 1$, with $\mathbf{C}_{ij} = 1$ for perfect correlation, $\mathbf{C}_{ij} = -1$ for perfect uncorrelation and $\mathbf{C}_{ij} = 0$ for uncorrelated components or assets for this case.

We can check the probability density function of the eigenvalues and eigenvectors to determine the level of unrandomness in companies correlations. First we have to calculate the amounts in the data: Q, λ_{min}, λ_{max} and compare with the theoretical values give by Eq. (3) and (4).

We have taken the high frequency prices of $M = 480$ companies of S&P500 each 30 min, since 09/09/2010 at 15:30:00 to 09/03/2011 21:30:00 (dd/mm/yyyy hh:mm:ss) for a total of $N = 1658$ for each asset, which makes $Q = 3.4541$.

2.3 Simulation

To compare the theoretical and return data, we have made simulations using Gaussian random numbers, this is called the Gaussian Orthogonal Ensemble (GOE), Edelman & Rao (2005). The GOE is defined on the space of real symmetric matrices which is very useful for the goal of this work: is invariant under orthogonal transformations, i.e. $Z \to Z' \equiv W^\dagger Z W$ where W is any real orthogonal matrix ($WW^\dagger = 1$), by this, conserve the joint probability $P(Z)dZ = P(Z')dZ'$, the second implication is that $\{Z_{ij} \mid i < j\}$ are statistically independent. For this, we build a correlation matrix $\mathbf{R}$ with the same dimension of our data: M series of N length with the same structure.

After compare the GOE and theoretical curve in Figure 1, we conclude that the range of eigenvalues is bigger than the theoretical curve because the finiteness of the data: N and $M \neq \infty$. The values are in Table 1.

	Theoretical	Simulated	Market
λ_{max}	2.3657	2.3662	149.5130
λ_{min}	0.2134	0.2167	0.0202

Table 1: Theoretical, simulated and market data

The highest eigenvalue obtained from $\mathbf{C}$ ($\lambda_{max}^{\mathbf{C}}$) is ≈ 60 times bigger than expected. In order to give a first approximation of the theory and the data, we set $\sigma = 0.16$ in Eq. (3) and (4) as is show in Figure 2, to show that the 16% of the variance of the data is represent by random part.

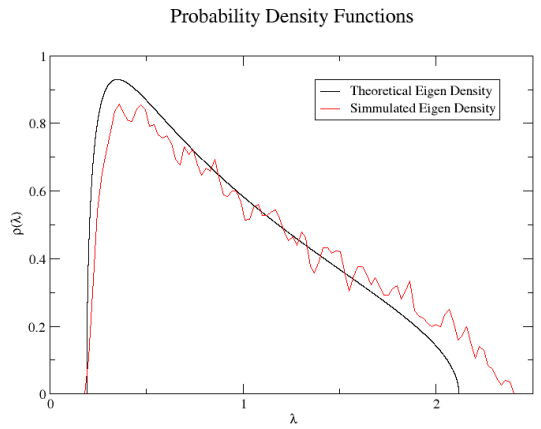


Figure 1: Probabilities Density Functions of Eigen Values of Theoretical (black line) and Simulated (red line).

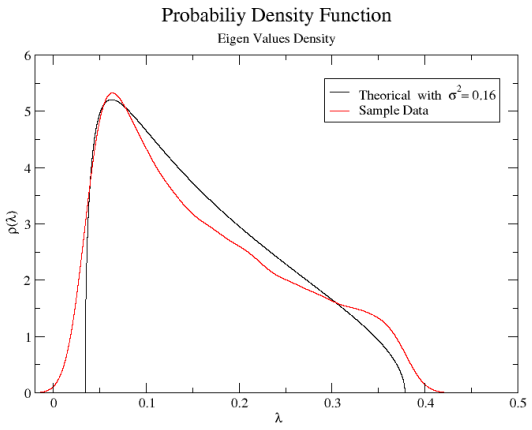


Figure 2: Probability Density Functions of theoretical eigenvalues density with $\sigma^2 = 0.16$ (black line) and sample eigenvalues density (red line).

2.4 PDF of Correlation Matrix

The Probability Distribution Function of the data $\mathbf{C}_{ij}$ and $\mathbf{R}_{ij}$ shows that $\langle \mathbf{C}_{ij} \rangle = 0.19$ (neglecting the cross terms $\mathbf{C}_{ii} = 0$) with standard deviation 0.27, against $\langle \mathbf{R}_{ij} \rangle = 0.0$ (also neglecting the cross terms $\mathbf{R}_{ii} = 0$) with standard deviation of 0.05. The Figure 3 shows bigger probability for $\mathbf{C}_{ij} > 0$ than GOE matrix, which reveal an important behaviour of the market: the positive correlations are more probable than negatives, at least for this period,[1] this is also watched if we calculate the mean of positive correlations and negative, $E[\mathbf{C}_{ij} \mid \mathbf{C}_{ij} > 0] = 0.24$ and $E[\mathbf{C}_{ij} \mid \mathbf{C}_{ij} < 0] = -0.10$. The absolute value is greater than positive as we expected. With this, we have find that $\mathbf{C}$ has different behaviour compared with $\mathbf{R}$ but also contain a random behaviour, showed

[1] Is also show in different relative works Plerou *et al.* (2002)

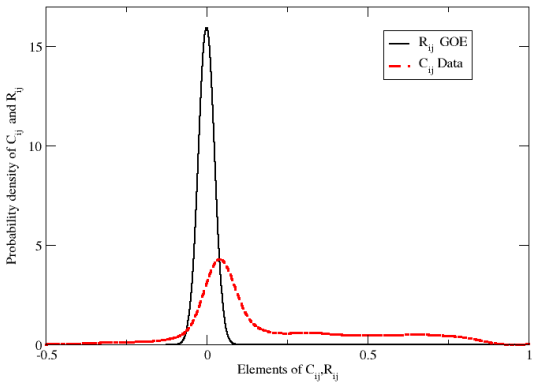


Figure 3: Probability Density Functions of the distributions of $\mathbf{R}_{ij}$ (solid line) and $\mathbf{C}_{ij}$ (dashed line).

in the picked part of the dashed line in Figure 3. Here, we can justify the eigenvalue analysis, by this, we expect find the different behaviors: random and nonrandom.

2.5 Inverse Participation Ratio

To separate the random and nonrandom part from $\mathbf{C}$ we will use the Inverse Participation Ratio concept (IPR). The IPR is used to measure the number of significant components of an eigenvector:

$$I^k \equiv \sum_{l=1}^{N} \left[u_i^k\right]^4 \tag{6}$$

where u_l^k is an eigenvector in position k with lth component. To understand the meaning of this quantity, let's see the limit cases: suppose that all component in the eigenventor are the same $u_l^k = 1/\sqrt{N}$, then $I^k = 1/N$ taking the inverse we have that has N components which contribute to the eigenvector, the other limit case is when only one component has value different from zero $u_1^k = 1$, applying again Eq. (6), we find that $I^k = 1$, which mean that only one component of the eigenvector contribute. We will use this concept, in order to analyze, which eigenstates have more or less significant contributors,for nonrandom eigenstates is expected that the numbers of *significant contributors*, or assets in this case, is lesser than random part, which is compared with the GOE in order to explore the difference carefully.

The range of the eigenvalues from $\mathbf{C}$ is bigger than eigenvalues from $\mathbf{R}$, as is show in Table 1, with bigger relative separation of the upper eigenvalues. The IPR in Figure 4 (top) and its inverse (button), show the dependence of the number of significant components for each eigenvector, the GOE is in the order of $\sim 10^{-2}$, which agree in order with N. $\mathbf{R}$

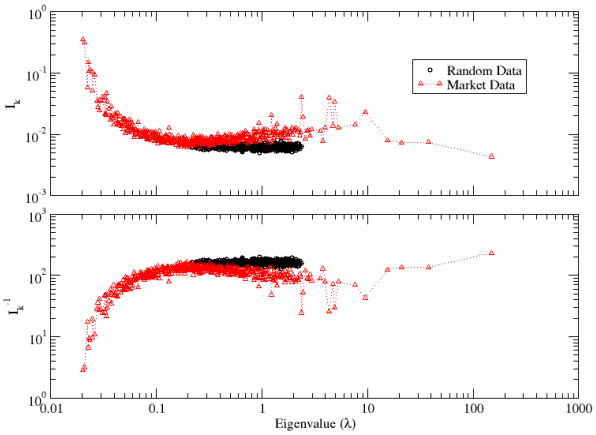


Figure 4: (top) Inverse Participation Ratio (IPR) as a function of eigenvalues of the Random data $\mathbf{R}$ (circle) and Market data C (triangle) $I_k(\lambda)$. (button) Number of significant contributors for each eigenvector as a function of eigenvalues $I_k^{-1}(\lambda)$.

has bounded eigenvalues as in theoretical predictions: Eq. (3) and (4). Comparing the eigenvalues of $\mathbf{C}$ with $\mathbf{R}$, we see clearly that the $\mathbf{C}$ eigenvalues have different significant IPR properties: lower than $\mathbf{R}(\lambda_{min})$ the IPR shows a low number of significant contributors $\sim 10^0$ and over than $\mathbf{R}(\lambda_{max})$ shows a larger number of significant contributors than GOE, as shows the Figure 4. With this, we have determined which eigenstates of $\mathbf{C}$ are outside of the random assumption and how many significant contributors, assets in this case, have each eigenstate.

2.6 Significance of Eigenstates

The largest eigenvalue $\mathbf{v}$ of $\mathbf{C}$, often called *the market mode*, is a important eigenstate and has a special meaning: this eigenstate is the one with more number of significant assets, it measure *something* that affect all assets, the *something* could be: news, special meetings, interest rate, etc. in generally depend of the data's period. To see the effect of this eigenstate in the market, is necessary use statistical decomposition, which is use by the Capital Asset Pricing Model (CAPM), Campbell *et al.* (1997):

$$Z(t) = \alpha + \beta Z_m(t) + \epsilon(t) \tag{7}$$

where Z_t is the return, α and β are adjustable parameters, $Z_m(t)$ is the mean return of eigenstate and $\epsilon(t)$ is the perturbation. In order to *separate* the eigenstate from the market, we have to find the perturbation ϵ_t for each asset, that balance the return of the market with the portfolio, in this case, the effect of the eigenvector of the largest eigenvalue $\mathbf{v}$, multiplied by the normalized return: $r^{LE}(t) = \sum_{i=1}^{M} r_i(t) v_i^{LE}$ which is common for all the assets. To solve the system, we have to find the effect of Eq. (7) for each asset i. Solving with ordinary least regression:

$$\beta_i = \frac{\sum_{j=1}^{N} r_i(j)(r^{LE}(t) - \mu_{LE})}{\sum_{j=1}^{t}(r^{LE}(t) - \mu_{LE})^2} \tag{8}$$

$$\alpha_i = -\beta_i \mu_{LE} \tag{9}$$

with $Z_m(t) = (1/N)\sum_{j=1}^{N} r^{LE}(j)$. Having this, we find from Eq. (7) $\epsilon_i(t) = Z_i(t) - \alpha_i + \beta_i Z_{im}(t)$, then calculate a new $\mathbf{C}_\epsilon$ from $\epsilon_i(t)$ without the effect of the largest eigenstate. As result we have $\langle \mathbf{C}_\epsilon \rangle = 0.03$ with standard deviation $= 0.10$ and $P(\mathbf{C}_\epsilon)$ as in shown in Figure 5.

The effect of the largest eigenstate has been take

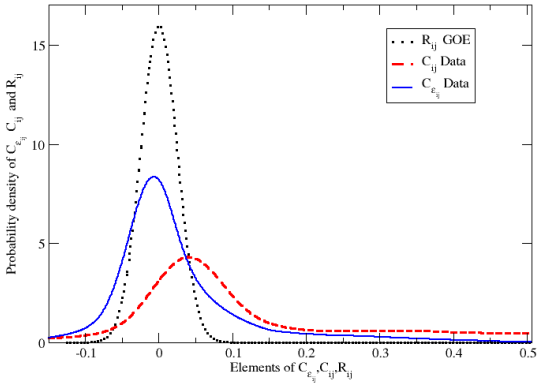


Figure 5: Probability Distribution Function (PDF) of the correlation matrix without the effect of the largest eigenvalue $\mathbf{C}_\epsilon$(solid line) compare with the PDF of a complete correlation matrix from data $\mathbf{C}$ (dashed line) and the PDF of random correlation matrix from GOE $\mathbf{R}$ (pointed line).

off in Figure 5, it uncenter and unweighted positively the correlation matrix, doing the $\mathbf{C}_\epsilon$ *more* random. Despite that the largest eigenstate was *take off* from $\mathbf{C}$, the $\mathbf{C}_\epsilon$ matrix has still some unrandom behaviour, by this, we will explore the following eigenstates. The meaning of the following eigenstates will be give by its number of significant assets contributors, Figure 4. The eigenstates greater than theoretical λ_{max} have group similar behaviuor, as capitalization, industry, locations, etc. The eigenstates lower than theoretical λ_{min}, e.g. the lowest eigenstate (eigenvector with the lowest eigenvalue), only has two number of significant assets contributors, which is represented by the largest absolute component of the eigenvector. For this case we have tabulate the results: Table 3. If we compare the assets in the eigenstates with some results from different works, Plerou *et al.* (2002), we confirm that *not always* the eigenstates after the biggest, separate the assets by sectors, we can confirm that they move for some common effect, which *sometimes* is the sector.

We center now, in the effect of the largest eigenstate, *the market mode*. For this, we have to eliminate the nonrandom part from the correlation matrix of the sample $\mathbf{C}$ which are represented by the eigenvalues less than the fifth biggest. The effect of the lowest is small compare with the others, by this we can neglect its contributions. To do this, we define a new matrix $\mathbf{\Lambda}'$ with the diagonal filled with the eigenvalues which we want conserve, in this case the random $\mathbf{\Lambda}'_{ii} = \{\lambda_1, \lambda_2, \cdots, 0, 0\}$. Done this, we transform $\mathbf{\Lambda}'$ with the basis of $\mathbf{C}$, we obtain the new $\mathbf{C}'$ without the effect of the eigenvalues neglected from $\mathbf{C}$ Plerou *et al.* (2002). The analysis will be applied to other $\mathbf{\Lambda}'$ this time with the eigenvalues eliminated before: $\mathbf{\Lambda}'_{ii} = \{0, 0, \cdots, \lambda_{M-1}, \lambda_M\}$, which will be the effect of the nonrandom part give by the correlation matrix, in others words, we take the more relevant eigenvalues which represent the biggest common effect in the dynamics of the market.

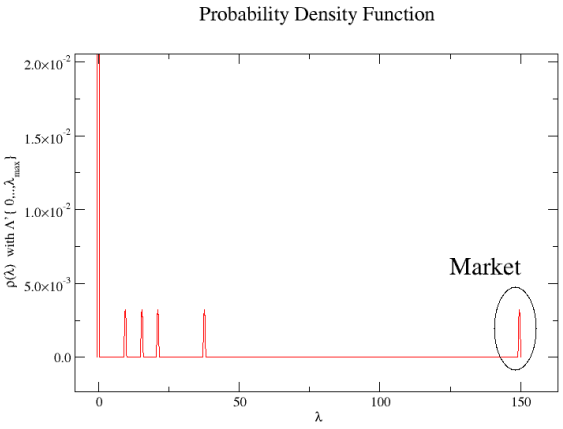


Figure 6: Probability Density Function of the eigenvalues of $\mathbf{C}' =$ basis of $\mathbf{C} \times \mathbf{\Lambda}'$ where $\mathbf{\Lambda}'$ is define as a diagonal matrix with $\mathbf{\Lambda}'_{ii} = \{0, 0, \cdots, \lambda_{M-1}, \lambda_M\}$

2.7 Application to Portfolio Theory

After measure structural group properties and study the effect of the eigenstates Market Data using RMT and IPR, we can use the theory of optimal portfolio selection Markowitz (1952), in order to measure the risk of the portfolio:

$$\Omega^2 = w^\dagger \mathbf{C} w \tag{10}$$

where Ω is the Risk of the portfolio, w is $m \times 1$ vector where each w_i the fraction of wealth invested on asset i normalized: $\sum_i^M w_i = 1$. $\mathbf{C}$ is the correlation matrix. With fixed return:

$$\mu_p = w\mu \tag{11}$$

where μ_p is the return of portfolio and μ is the return of assets. We have to solve a minimize problem: with a fixed portfolio return μ_p and with a normalized portfolio $\sum_i^M w_i = 1$, we have to find the family of portfolios which give the minimum risk Markowitz (1952). The solution is easily find with Lagrange multipliers and the two constraints:

$$w(\mu_p) = \frac{1}{BC - A^2}(B\mathbf{C}^{-1}1 - A\mathbf{C}^{-1}\mu + \mu_p(C\mathbf{C}^{-1}\mu - A\mathbf{C}1)) \tag{12}$$

with:

- $1 = (1, \cdots 1)^\dagger$
- $A = \mu^\dagger \mathbf{C} 1$
- $B = \mu^\dagger \mathbf{C} \mu$
- $A = \mu^\dagger \mathbf{C} 1$

We separate the data in two equal periods first and second (1 and 2), finding the family of optimal portfolios with the matrix correlation using $\mathbf{C}$, is assumed that we have perfect knowledge of the future by this, we take the return of second period μ_2 as the predicted return, the risk is calculated using $\mathbf{C}_1$, the correlation matrix from first period ($\mathbf{C}_1$), we will called the *predicted risk*. Also we calculate the *realized risk*, which consist in take the same family of portfolios but the risk is calculated with the correlation matrix of second period $\mathbf{C}_2$ Eq. (10). We find the *efficient portfolio frontier* of risk and return Figure 7 and 8 for predicted and realized risk (Ω_p, Ω_r). We found that the relative

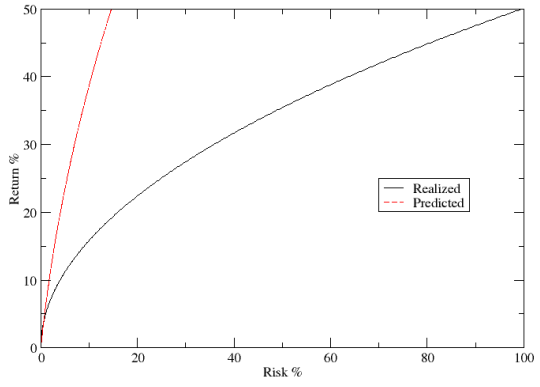


Figure 7: Relation between a given Return of mean-variance portfolios and Risk calculated using the correlation matrix $\mathbf{C}_1$. (dashed line) Relation measuring risk from $\mathbf{C}_1$. (solid line) Relation measuring risk from $\mathbf{C}_2$.

error is:

$$\frac{\Omega_r^2 - \Omega_p^2}{\Omega_p^2} \approx 150\% \tag{13}$$

Now, we will see the effect of use the RMT to *clean* the correlation matrix $\mathbf{C}$. With the method defined in

before subsection (2.6), we will eliminate the random part from $\mathbf{C}$, we will called the cleaned correlation matrix $\mathbf{C}'$. In order to compare we will find an optimal family of portfolio with $\mathbf{C}'$ and as the before part, we will calculated the risk with both correlation matrix periods but this time cleaned (Ω_p, Ω_r) We found an

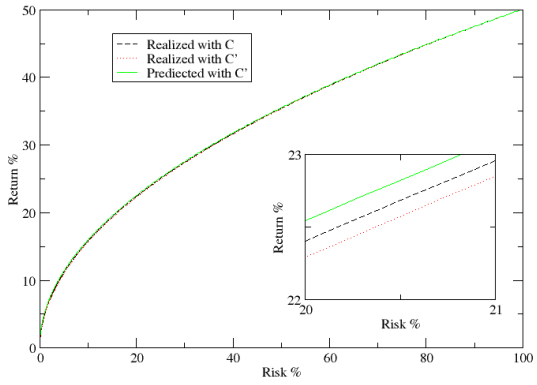


Figure 8: Relation between a given Return of mean-variance portfolios and Risk calculated using the cleaned correlation matrix $\mathbf{C}'$. (dashed line) Relation measuring risk from $\mathbf{C}_1$. (doted line) Relation measuring risk from $\mathbf{C}'_1$. (solid line) Relation measuring risk from $\mathbf{C}'_2$.

error:

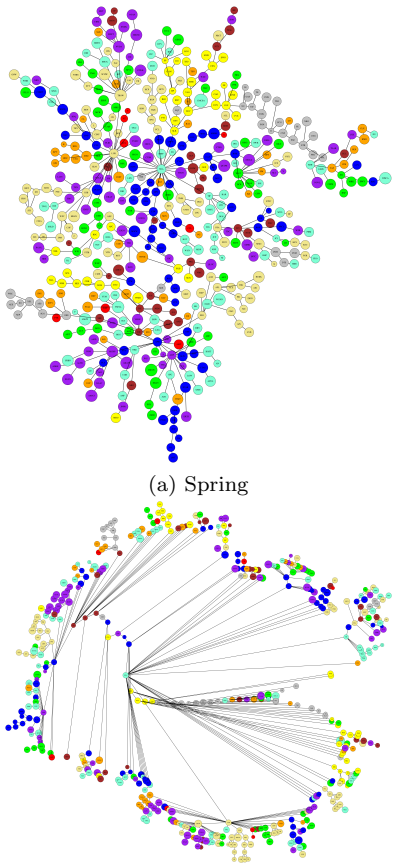
$$\frac{\Omega_r^2 - \Omega_p^2}{\Omega_p^2} \approx 10\% \tag{14}$$

which is lower than the calculated with non cleaned correlation matrix.

3 Complex Network Approach

Group properties, has been well studied in the previous section, now we use the the complex network approach, in order to measure different and important group properties. The market can be modeled as a weighted and undirected network, defining the metric as $M = 1 - \mathbf{C}_{ij}, \forall \mathbf{C}_{ij} > 0$, at first we have a dense network, all component connect with the others, that can be simplify with several algorithms as the Minimum Spanning Tree (MST), which consist in find an acyclic subset of companies that connects with all of the others with a minimized metric, that is, find the biggest possible correlation without generate a close loop. With this, we have simplify the relations between stocks, as result, we can build a MST network shows in Figure 9, where the colors are given by the sector of each company, Table 2.

As result we expect that the network separate the companies by sectors Aste *et al.* (2010), after look for clusters, we found that the companies are group in different subsets, which mean, that the companies are



(a) Spring

(b) Radial

Figure 9: Plot of Graph.

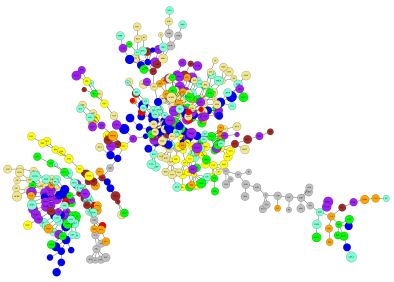
correlated in general with any other, that could be explain by location, same clients, etc. as show the Figure 10.

The network model reveals and important property in the market, we see in Figure 9, that some companies are *more connect* than others, this property is called the degree of a vertex, which in the network after the MST, is the company with more number of correlation that *survive* to simplification, that is, the nearest common company with the metric before defined. We found, in our data, these companies are SNA and LUK with 29 and 25 respectively.

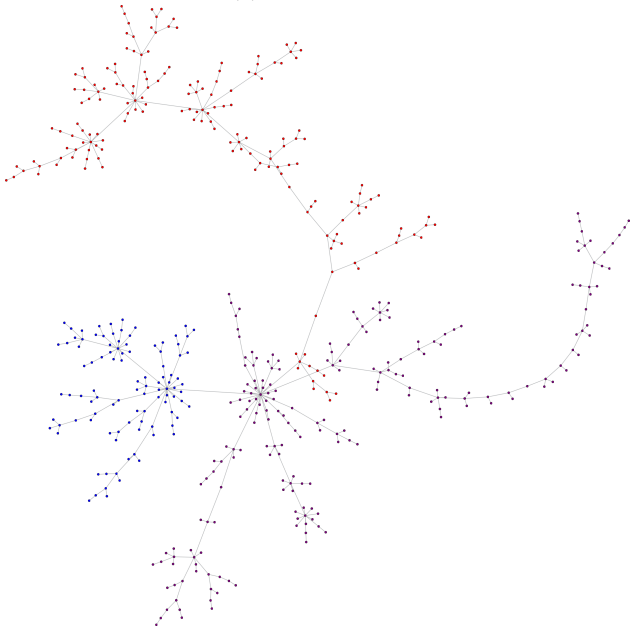
The colors indicates the sectors of each company.

Sector	Color	N. of Com.
Telecommunications Services	Red	8
Industrial	Blue	62
Energy	Yellow	39
Utilities	Gray	35
Consumer Staples	Orange	41
Health Care	Green	51
Materials	Brown	30
Information Technology	Purple	73
Consumer Discretionary	Aquamarine	79
Financial	Khaki	82

Table 2: Financial Sectors



(a) Spring cluster



(b) Cluster

Figure 10: Plot of Graph by Clusters.

4 Conclusions

In this work we have study the group behaviour of 480 companies in the S&P500 for seven months. Using the Random Matrix Theory, we have show that the correlation matrix produced by the 480 companies is in most part random. We have eliminate the random part to show the effect in the correlation matrix and then we have eliminated the nonrandom part in order to apply the Markowitz theory to calculated the family of optimal portfolios and measure the efficient frontier of wealth and risk, with very good results. In the second part, we have define a network from the Minimum Spanning Tree algorithm apply to the relations in the correlation matrix, as consequence we have discovered certain relevant companies with more connections that many others, opening with this, to possible financial portfolio strategies based in a non-risk assets, where the behaviour of such companies could define the others with less connections, to analyze this more deeply, we could implement a epidemic models to determine the dynamics of this effect.

The cluster effect in a group of companies, could be determine in several ways: first, using the Inverse Participation Ratio to see how many and which companies are affected in each eigenstate, being the importance of such eigenstate proportional of its eigenvalue e.g. some recent news, can *clusterize* the market. Second, using the network approach, we can see how the companies are correlated, in this sense, we can define a new portfolio with the more correlated assets, using the shortest path concept, in order to determine the largest more correlated portfolio, using the risk free asset as we mentioned before.

For future work, we expect analyze more companies in a larger period, in order to measure the dynamics of the properties exposed here and look for others that we expect and could not find, debt to the short data term, as the autocorrelations and leverage effect, between companies and index.

References

Aste, T., Shaw, W. & Matteo, T. D. 2010 Correlation structure and dynamics in volatile markets. *New Journal of Physics* **12** (8), 085009.

Bachelier, L. 1900 Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure Sér.* **3** (17), 21–86.

Campbell, J., Lo, A. & MacKinlay, A. 1997 *The econometrics of financial markets*, 2nd edn. Princeton, NJ: Princeton Univ. Press.

Chakraborti, A., Muni Toke, I., Patriarca, M. & Abergel, F. 2009 Econophysics: Empirical facts and agent-based models. *Quantitative Finance preprints* .

Edelman, A. & Rao, N. R. 2005 Random matrix theory. *Acta Numer.* **14**, 233–297.

Laloux, L., Cizeau, P., Bouchaud, J.-P. & Potters, M. 1999 Random matrix theory and financial correlations. Science & Finance (CFM) working paper archive 500053. Science & Finance, Capital Fund Management.

Markowitz, H. 1952 Portfolio selection. *Journal of Finance* **7** (1), 77–91.

Miller, S. J. & Takloo-Bighash, R. 2007 *Introduction to Random Matrix Theory from An Invitation to Modern Number Theory*. PRINCETON UNIVERSITY PRESS.

Osborne, M. F. M. 1959 Brownian motion in the stock market. *Operations Research* **7** (2), 145–73.

Osborne, M. F. M. & Murphy Jr., J. E. 1984 Financial analogs of physical Brownian motion, as illustrated by earnings. *Financial Review* **19** (2), 153–172.

Eigenstate 0	
SE	Energy
LOW	Consumer Discretionary
CAH	Health Care
TER	Information Technology
ESRX	Health Care
CSC	Information Technology
HNZ	Consumer Staples
SCG	Utilities
DO	Energy
ROK	Industrials
Eigenstate 1	
NBL	Energy
BEN	Financials
CAH	Health Care
QLGC	Information Technology
SCG	Utilities
SE	Energy
JCI	Consumer Discretionary
CSC	Information Technology
ESRX	Health Care
TXN	Information Technology
Eigenstate 2	
CSC	Information Technology
NEM	Materials
CVH	Health Care
CAH	Health Care
COH	Consumer Discretionary
UNM	Financials
BEN	Financials
TER	Information Technology
TSO	Energy
EIX	Utilities
Eigenstate 3	
LOW	Consumer Discretionary
BEN	Financials
SCG	Utilities
TIE	Materials
CAH	Health Care
CTSH	Information Technology
UNM	Financials
NSC	Industrials
CSC	Information Technology
ESRX	Health Care

Table 3: Biggest Companies and sector in each eigenstate.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T. & Stanley, H. E. 2002 Random matrix approach to cross correlations in financial data. *Phys. Rev. E* **65** (6), 066126.