Investigation of Cluster Structure in the London Stock Exchange

Ricardo Coelho, Przemyslaw Repetowicz, Stefan Hutzler and Peter Richmond
School of Physics, Trinity College, University of Dublin

science foundation ireland
fondúireacht eolaíochta éireann

INTRODUCTION

Since the last decade, many physicists have applied techniques of statistical physics to the study of economics and finance [1, 2, 3]. A mathematical theory of ensembles of random matrices that are subject to certain invariance properties (Random Matrix Theory [4]) that first served as a tool for describing interactions in Nuclear Physics has provided inspiration for the investigation of correlations in multi-variate Time Series. Our objective is to model multi-variate series of stock returns as simple random moving average models and, in addition to that, estimate the underlying correlation matrix in the series. First we analyze a portfolio of stocks from FTSE100 by computing the sample correlation matrix, we define a metric in the space of time series and we construct a Minimal Spanning Tree that shows us how different companies are linked with each other. Then we generate uncorrelated time series, construct their MSTs and compare the results to the former ones. Finally we fit the FTSE series to the FTSE index benchmark by mean square regression and we compare the MST structure to previous results.

1 Taxonomy of the Market

Our main point is to detect similarities between stocks. So, comparing the time series of log-returns, we study the sample correlation coefficient between pairs of stocks. From these correlations we can compute a distance, for each pair, that will help us to construct a network with links between stocks.

1.1 Correlations

The sample correlation coefficient is a relation between the time series of different stocks:

$$\rho_{ij}(t) = \frac{< R_i R_j > - < R_i >< R_j >}{\sqrt{\left(< R_i^2 > - < R_i >^2\right)\left(< R_j^2 > - < R_j >^2\right)}}$$

where $R_i(t) = \ln P_i(t) - \ln P_i(t-1)$ is the logarithmic return and $P_i(t)$ the daily closure price of stock i at time t. The sample correlation coefficients can vary between $-1 \le \rho_{ij} \le 1$, where -1 means completely anticorrelated and $+1$ completely correlated. When $\rho_{ij} = 0$ the stocks i and j are uncorrelated. This coefficients form a symmetric $N \times N$ matrix where the diagonal elements are equal to 1 ($\rho_{ii} = 1$).

1.2 Distances

A metric distance, introduced by Mantegna [5], is very useful to our study of the taxonomy of market. This distance is determined from the Euclidean distance between vectors, that are the time series of logarithm returns $\vec{R_i}$. The log-returns are assumed to be independent, identicaly distributed random variables with mean zero and a finite variance. In this case the sample correlation coefficient is an unbiased estimator of the population correlation coefficient. We do not investigate problems related to estimators of correlation coefficients in multi-variate distributions with infinite second moment. We normalize each time series:

$$\tilde{R}_i = \frac{R_i - < R_i >}{\sigma_{R_i}}$$

The distance is represented by $d_{ij} = |\tilde{R}_i - \tilde{R}_j|$. If we consider:

$$d_{ij}^2 = |\tilde{R}_i - \tilde{R}_j|^2 = |\tilde{R}_i|^2 + |\tilde{R}_j|^2 - 2\tilde{R}_i \cdot \tilde{R}_j = 2 - 2\rho_{ij}$$

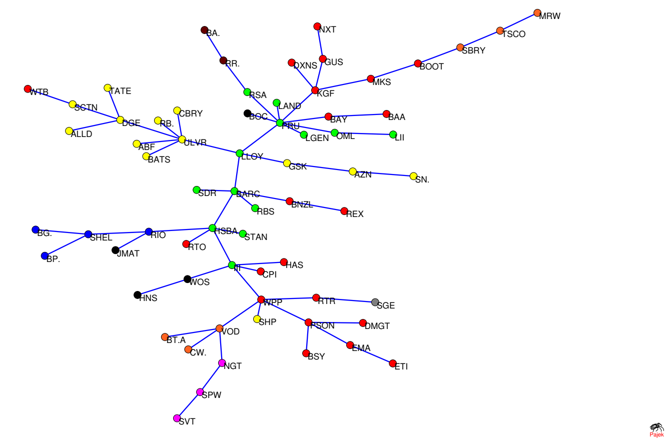
where $\rho_{ij} = \tilde{R}_i \cdot \tilde{R}_j$, we have a relation between the distance of two stocks and their sample correlation coefficient:

$$d_{ij} = \sqrt{2(1-\rho_{ij})}$$

This distance has values between $0 \le d_{ij} \le 2$. Small values mean strong correlations between stocks. Each distance is part of a matrix of distances with the same properties as the correlation matrix.

1.3 Minimal Spanning Trees

The distance matrix is now used to construct a network with the essential information of the market. This network is a Minimal Spanning Tree (MST), where nodes and edges are stocks and distances between them, respectively. The MST has $N-1$ edges, connecting N stocks, such that the sum of all distances (tree length) is minimum. We use the Prim's algorithm [6] to compute the MST. The portfolio investigated here is composed of 67 stocks from FTSE100 with time series length of about 9 years (2322 days) from August 1996 until June 2005. Figure 1 represents the MST.



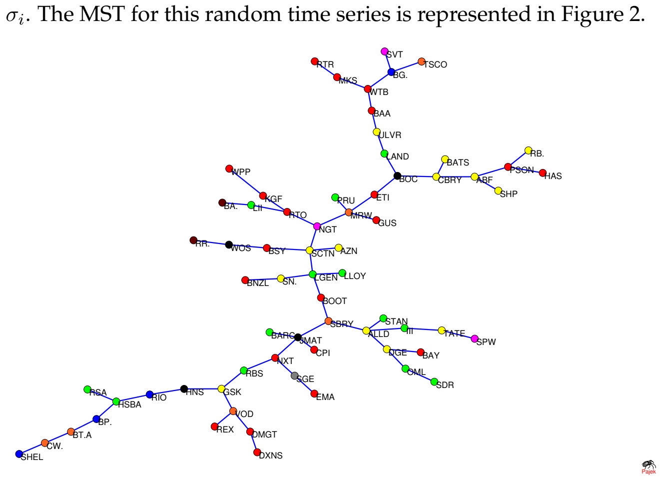
Our portfolio has 9 different Economic Groups that we represent with a different color: Resources (blue), Basic Industries (black), General Industrials (brown), Non-cyclical Consumer Goods (yellow), Cyclical Services (red), Non-cyclical Services (orange), Utilities (magenta), Financials (green) and Information Technology (gray). As can be seen in the MST, different stocks from the same Economic Group cluster together and the Financial Group (green), that has all stocks linked together, works as the backbone of this network. The distribution of number of neighbours of stocks show that some stocks have a few links but others have up to 8 links.

2 Random Trees

If we model the returns as random numbers from a specific distribution, we can compute the correlations, distances and trees for this random series. Our first approach was to consider returns as random variables from a Gaussian distribution [7]. So, using the mean value of each real time serie and the specific variance, we compute our random series:

$$R_i(t) = \mu_i + \epsilon_i(t)$$

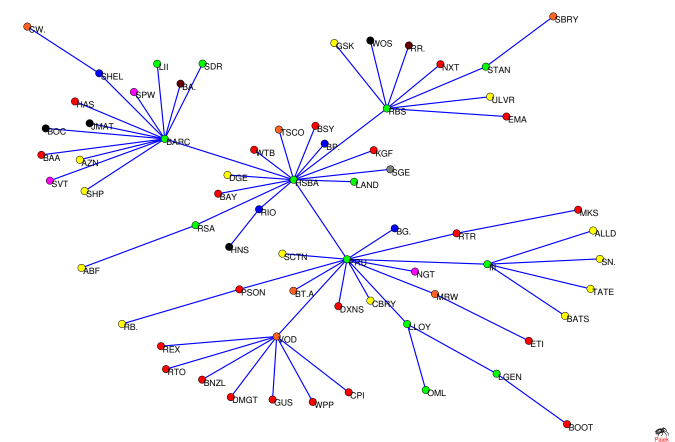
where μ_i is the mean and $\epsilon_i(t)$ is the stochastic variable from a Gaussian distribution with variance σ_i. The MST for this random time series is represented in Figure 2.



This MST doesn't show any cluster, the stocks are distributed randomly in the network and there is not any stock with more than 4 links. To change this, and trying to create random time series with more real characteristics we introduce a control term (the return of FTSE Index):

$$R_i(t) = \alpha_i + \beta_i R_m(t) + \epsilon_i(t)$$

where α_i and β_i are real parameters (estimated by the least square method), $R_m(t)$ is the market factor (return of FTSE Index) and $\epsilon_i(t)$ is the stochastic variable from a Gaussian distribution with variance σ_i. This model is known as the Market Model [8]. The MST for random time series created with the Market Model is represented in Figure 3.



This network is completely different from the other random one, here we can see that the stocks from the Financial Group (green), that are all linked together, still work as the backbone of the network. However, the presence of 6 nodes with up to 13 links does not reflect the topology shown in real data.

3 Conclusions

- The FTSE time series cannot be approximated well by uncorrelated multivariate Gaussian distributions.
- The market factor (Capital Asset) does not provide a satisfactory description of the series in terms of topology of the MST, although it reproduces some clustering of related stocks.
- Time series with infinite variance cannot be studied with this technique. New tools need to be developed.
- This work provides inspirations for novel techniques of parameter estimation in multi-variate time series with finite variance.
- For the investor, MST provide a useful tool for portfolio optimization.

Acknowlegments

Research was funded by Science Foundation Ireland, SFI BRG04/BR/P0251.

References

[1] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge (2001)

[2] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge (2003)

[3] P. Richmond, P. Repetowicz and S. Hutzler, "Dynamics of money and income distributions II", in *Econophysics of Wealth Distributions, Econophys-Kolkata I*", A. Chatterjee, S. Yarlagadda and B.K. Chakrabarti (Eds.), Springer-Verlag Italia, 120-125.

[4] M. L. Mehta, *Random Matrices*. Academic Press (1991)

[5] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999)

[6] R. C. Prim, Bell System Technical Journal **36**, 1389 (1957)

[7] G. Bonanno, G. Caldarelli, F. Lillo and R. N. Mantegna, Phys.Rev. E **68**, 046130 (2003)

[8] Y. J. Campbell, A. W. Lo and A. C. Mackinlay, *The Econometrics of Financial Markets*. Princeton University Press, Princeton (1997)