OCS

Correlation, hierarchical clustering and correlation based networks in financial markets

Rosario Nunzio Mantegna

Palermo University, Italy

Observatory of Complex Systems

OCS

From today Financial Times

Global markets

Americas | Europe | Asia-Pacific

FTSE Eurofirst 300 -46.39 / -4.94%

Index 5-day change

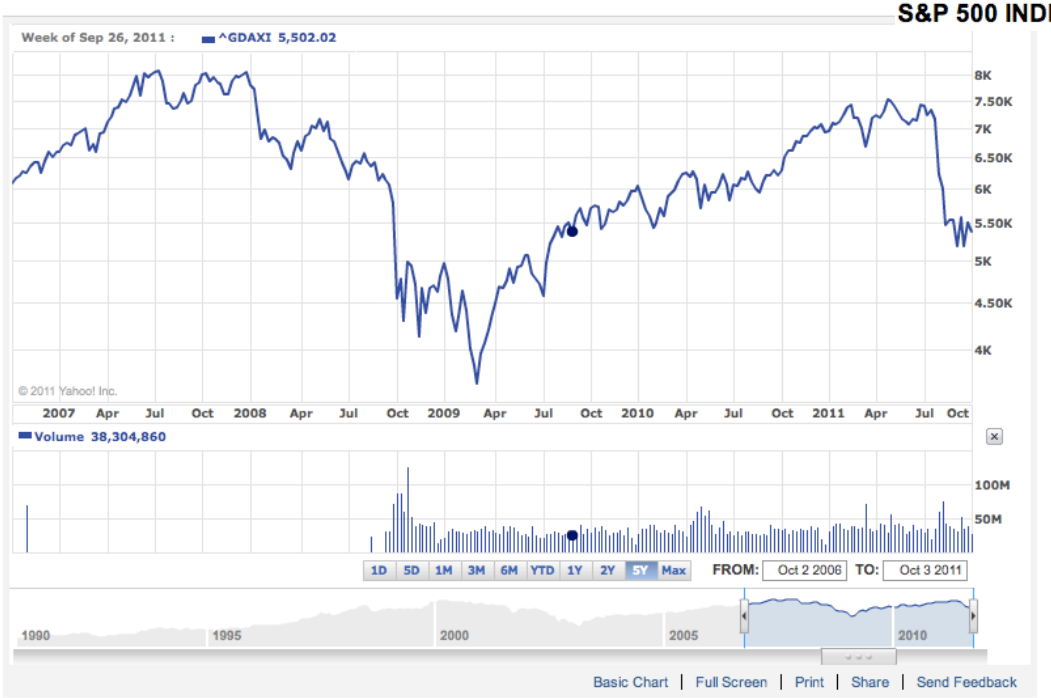
Index	Last	Change	Volume	Low 52-week High
FTSE Eurofirst 300 Open	891.99	-20.27 -2.22%	75.82m	
FTSE 100 Open	4,962	-113.61 -2.24%	249.74m	
CAC 40 Open	2,854	-72.98 -2.49%	61.29m	
Xetra DAX Open	5,219	-157.53 -2.93%	60.67m	

As of Oct 04 2011 10:06 BST. United Kingdom details

European equities news

OCS

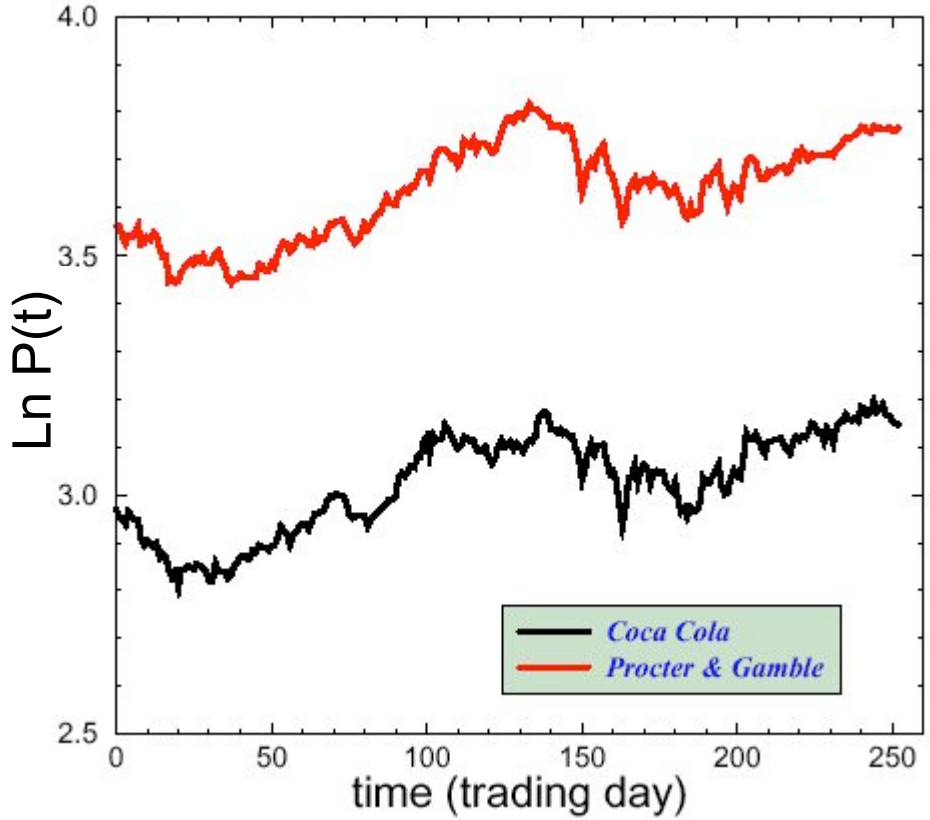
S&P 500 INDEX,RTH (SNP: ^GSPC)



DAX (XETRA: ^GDAXI)

OCS

Cross-correlation between pairs of stock returns are well-known



$$r_i(t) \equiv \ln P_i(t) - \ln P_i(t-1)$$

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\langle r_i^2 - \langle r_i \rangle^2 \rangle \langle r_j^2 - \langle r_j \rangle^2 \rangle}}$$

They may be quantified by the correlation coefficient ρ_{ij}

OCS

Overview

How to quantify and model information present in a correlation matrix?

- Describe the structure of empirical correlation matrices;

- Validate the statistical robustness of the detected correlation matrices with methods from Random Matrix Theory;

- Model hierarchies of a complex system in terms of hierarchical trees and correlation based networks;

-Validate the statistical robustness of correlation based networks with bootstrap procedures;

- Provide a factor model which is able to describe a system with a nested hierarchy;

- Provide a quantitative tool to compare different filtering methods.

OCS

A basic approach: Principal Component Analysis

A first objective of principal component analysis is to determine the standardized linear combination of the original variables which has maximal variance.

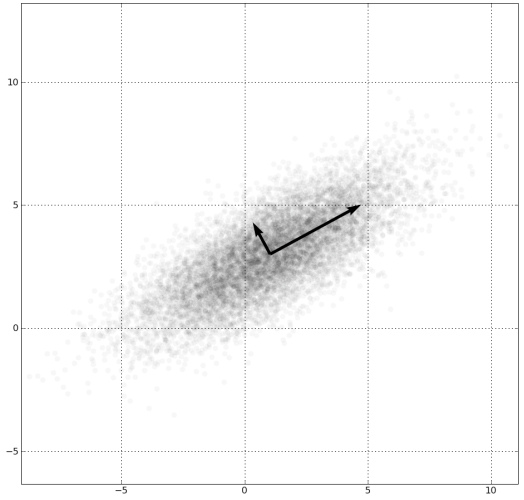
Principal component analysis looks for a few linear combinations which can be used to summarize data, losing in the process as little information as possible.

OCS

$\mathbf{C}$ = correlation matrix

$\mathbf{D}$ = diagonal matrix of eigenvalues of $\mathbf{C}$: $\mathrm{diag}[\lambda_1, \lambda_2, ..., \lambda_N]$

$\mathbf{V}$ = orthogonal matrix of eigenvectors of $\mathbf{C}$



Given a set of points in Euclidean space, the first principal component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the distances of the points from the line.

The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted out from the points.

OCS

Each eigenvalue of the correlation matrix is proportional to the portion of the "variance" (more correctly of the sum of the squared distances of the points from their multidimensional mean) that is correlated with each eigenvector. The sum of all the eigenvalues is equal to the sum of the squared distances of the points from their multidimensional mean.

OCS

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.413	0.518	0.543	0.529	0.341	0.271	0.231	0.412	0.294
IBM		1	0.471	0.537	0.617	0.552	0.298	0.475	0.373	0.270
BAC			1	0.547	0.591	0.400	0.258	0.349	0.370	0.276
AXP				1	0.664	0.422	0.347	0.351	0.414	0.269
MER					1	0.533	0.344	0.462	0.440	0.318
TXN						1	0.305	0.582	0.355	0.245
SLB							1	0.193	0.533	0.592
MOT								1	0.258	0.166
RD									1	0.590
OXY										1

The eingevalue spectrum of this correlation matrix is

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
4.72	1.40	0.96	0.52	0.48	0.45	0.42	0.39	0.35	0.30

OCS

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
	4.72	1.40	0.96	0.52	0.48	0.45	0.42	0.39	0.35	0.30
Percent of variance	47.2	14.0	9.6	5.2	4.8	4.5	4.2	3.9	3.5	3.0
Total variance explained	47.2	61.2	70.8	76.0	80.8	85.3	89.5	93.4	96.9	99.9

The most important principal components are

$$y_1 = 0.31x_1 + 0.34x_2 + 0.33x_3 + 0.35x_4 + 0.38x_5 + 0.32x_6 + 0.27x_7 + 0.27x_8 + 0.31x_9 + 0.26x_{10}$$
$$y_2 = 0.03x_1 + 0.22x_2 + 0.15x_3 + 0.13x_4 + 0.17x_5 + 0.23x_6 - 0.50x_7 + 0.32x_8 - 0.40x_9 - 0.56x_{10}$$
$$y_3 = 0.50x_1 - 0.12x_2 + 0.33x_3 + 0.30x_4 + 0.12x_5 - 0.43x_6 - 0.18x_7 - 0.54x_8 - 0.03x_9 - 0.12x_{10}$$
$$y_4 = 0.37x_1 - 0.42x_2 + 0.33x_3 - 0.43x_4 - 0.20x_5 + 0.11x_6 - 0.37x_7 + 0.33x_8 + 0.25x_9 + 0.17x_{10}$$
$$y_5 = \ldots\ldots$$

AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY

OCS

Selection of the most relevant principal components:

(i) variance threshold (ex: 90%);
(ii) “scree graph” (percentage of variance explained by each eigenvalue (due to Cattel,1966);
(iii) exclude those principal components whose eigenvalues are less than average, i.e. less than one if a correlation matrix is used (Kaiser).

OCS

How can we estimate and filter the information present in a correlation matrix estimated from a finite number of records?

Example: what is the empirical structure of the return correlation matrix of a large portfolio of stocks?

OCS

Cross Correlation

N data series of length T

$$r_i(t_j), \quad j = 1,\ldots,T; \quad i = 1,\ldots,N$$

Example:
Log-return of stock price

$$r_i(t) \equiv \ln P_i(t) - \ln P_i(t-\tau)$$

Pearson's correlation coefficient:

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\left\langle r_i^2 - \langle r_i \rangle^2 \right\rangle \left\langle r_j^2 - \langle r_j \rangle^2 \right\rangle}}$$

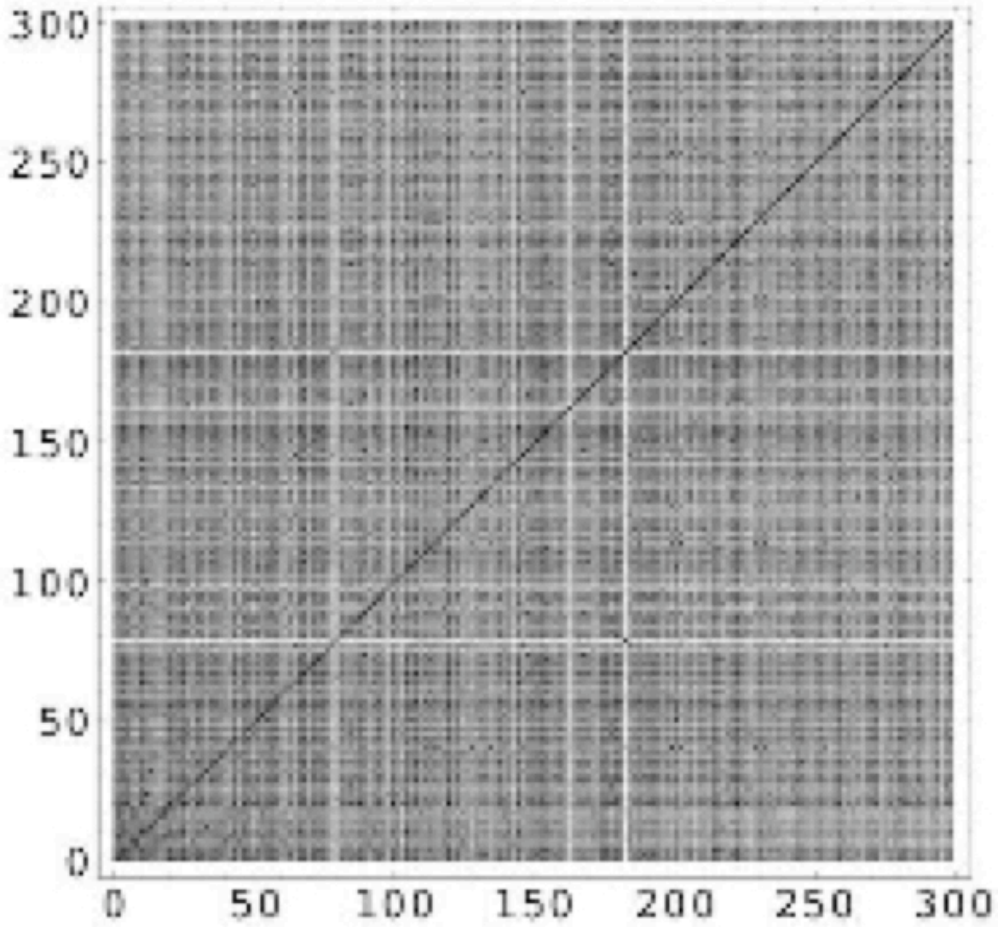
Other correlation estimators:

-Fourier estimator
-Maximum Likelihood correlation estimator
-Hayashi-Yoshida estimator
- .....

Correlation Matrix

$$C = \left(\rho_{ij}\right)$$

Grayscale representation of the return correlation matrix of a portfolio of stocks



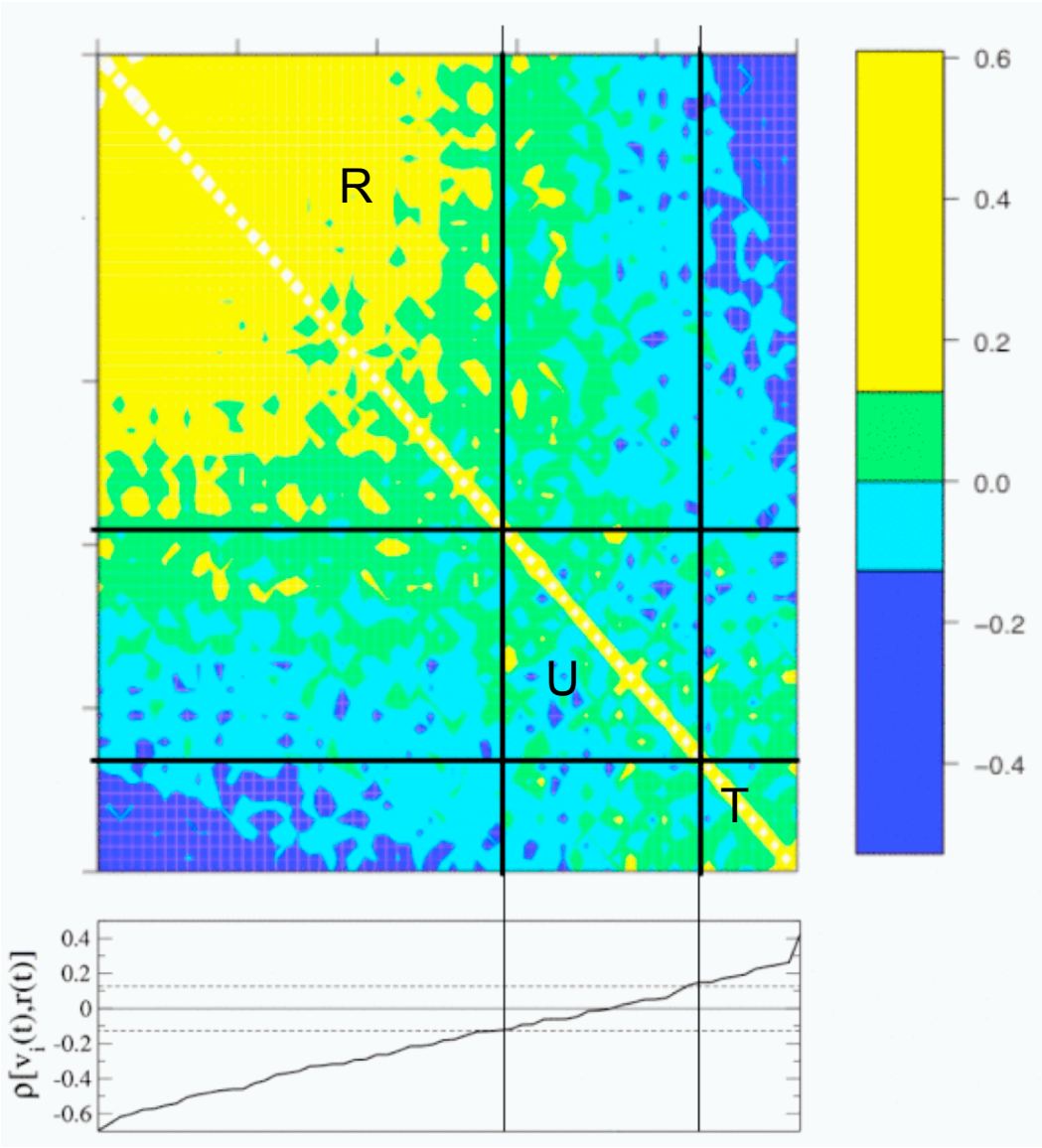
n(n-1)/2 distinct correlation coefficients

300 stocks traded at the US equity markets in 2001-2003

OCS

Inventory variation correlation matrix obtained by sorting the MMs in the rows and columns according to their correlation of inventory variation with price return

BBVA 2003



OCS

Statistical reliability of cross correlation coefficients

N T data $\longrightarrow \sim N^2$ correlation coefficients:

Statistical uncertainty is unavoidably associated with the estimation of the correlation coefficient obtained from a finite number of records.

It is therefore important to device methods to

- Filter statistically reliable information;

- Quantitatively assess the stability of the filtered information;

- Model the filtered information.

OCS

How to analyze the complexity of a correlation matrix?

Random Matrix Theory

Clustering e.g. Hierarchical Clustering
Super Paramagnetic Clustering
Maximum Likelihood Clustering
Sorting Point Into Neighbors

Correlation Based Networks e.g. Minimum Spanning Tree (MST)
Planar Maximally Filtered Graph (PMFG)
Average Linkage Minimum Spanning Tree

M. Tumminello, F. Lillo, R.N. Mantegna, Correlation, hierarchies, and networks in financial markets, Journal of Economic Behavior & Organization (2010).

OCS

Random Matrix Theory

Random Matrix Theory allows to quantify
the degree of statistical uncertainty observed
in the determination of a correlation coefficient
matrix done with a finite number of records
(under some simplifying assumptions).

Madan Lal Mehta, Random matrix theory, Academic press 1990

OCS

RMT for Gaussian variables

Suppose that N elements are described by N time series of length T. They are independent Gaussian random variables with zero mean and variance σ^2.

The correlation matrix of this set of variables in the limit T -> ∞ is the identity matrix, which has associated a spectrum of eigenvalues which is composed of N delta functions at $\lambda=1$

OCS

The limits of validity

However, due to statistical uncertainty for finite values of T what it is really observed for the eigenvalue spectrum of the correlation coefficient matrix is a spectrum characterized by eigenvalues with values located between λ_{min} and λ_{max}

In the limit T, N -> ∞ with a fixed ratio Q=T/N $\geq$1 random matrix theory provides the probability density of the eigenvalues.

OCS

PDF of eigenvalues

The probability density is

$$\rho(\lambda) = \frac{T}{2\pi\sigma^2\lambda}\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}$$

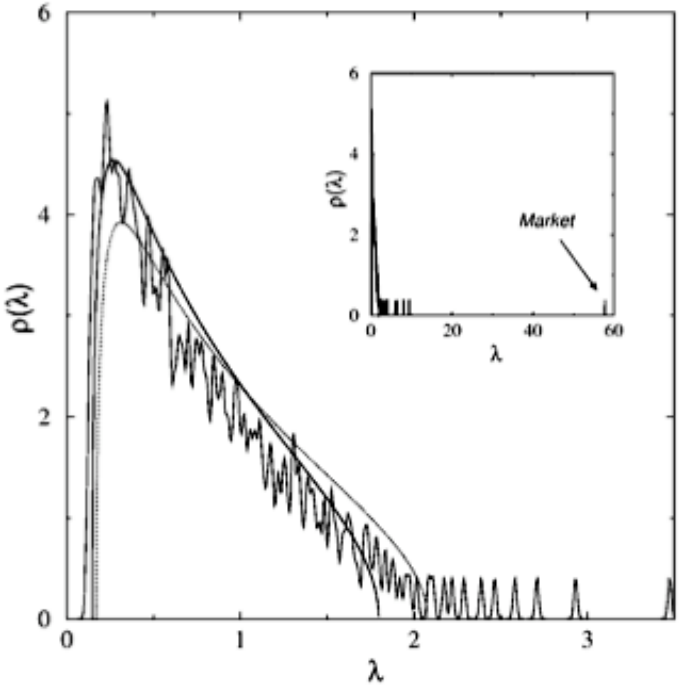
where

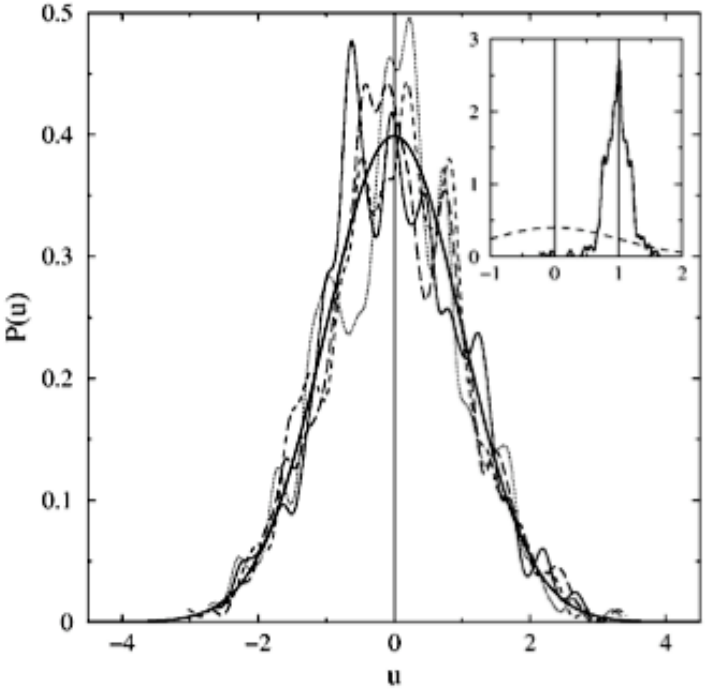
$$\lambda_{\min}^{\max} = \sigma^2\left(1 + 1/Q \pm 2\sqrt{1/Q}\right)$$

with Q=T/N

OCS

Random Matrix Theory of financial correlation





$$\rho(\lambda) = \frac{T}{2\pi\lambda\sigma^2}\sqrt{(\lambda_{\max}-\lambda)(\lambda-\lambda_{\min})}, \quad \text{where} \quad \lambda_{\min}^{\max} = \sigma^2\left(1+1/Q \pm 2\sqrt{1/Q}\right), \quad \sigma^2 = 1-\frac{\lambda_1}{N}, \quad \text{and} \quad Q = \frac{T}{N}.$$

L. Laloux, P. Cizeau, J.-P. Bouchaud & M. Potters, *Phys. Rev. Lett.* **83**, 1468 (1999).
V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, *Phys. Rev. Lett.* **83**, 1471 (1999).

OCS

Characteristics of the eigenvalue spectrum

The spectrum of a typical portfolio can be divided in three classes of eigenvalues:

1) The largest eigenvalue describes the common behavior of stocks (what is called “the market”). It is incompatible with the random matrix theory of random variables.

2) A fraction of 5% of the eigenvalues is also incompatible with the random matrix theory because eigenvalues fall outside $]\lambda_{min}, \lambda_{max}[$

OCS

The low part of the spectrum

3) The remaining eigenvalues assume values between λ_{min}, λ_{max} and therefore one cannot say whether the eigenspace, which is corresponding to these eigenvalues, contains information or not.

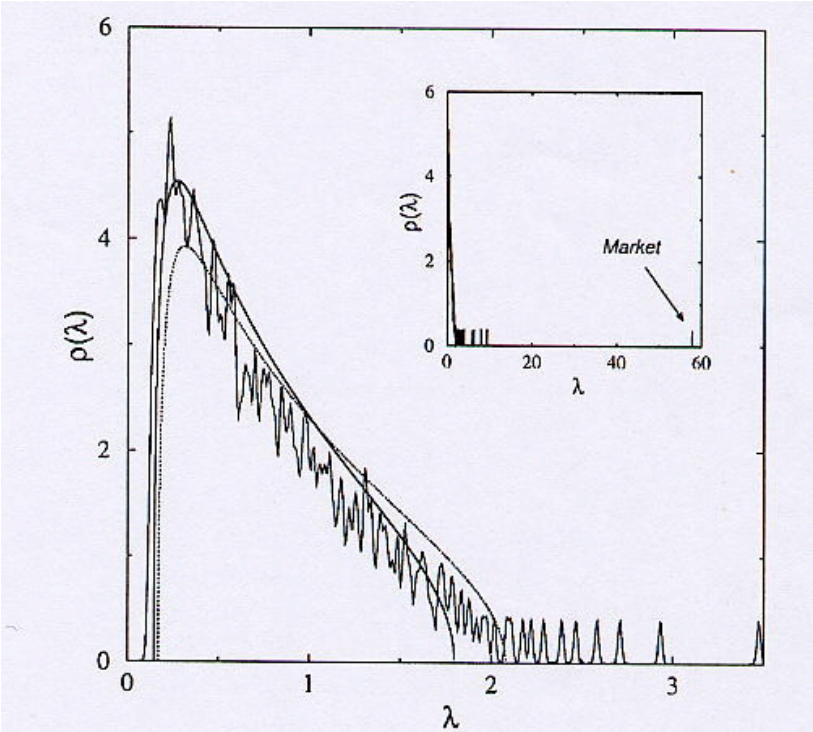


FIG. 1. Smoothed density of the eigenvalues of **C**, where the correlation matrix **C** is extracted from $N = 406$ assets of the S&P 500 during the years 1991–1996. For comparison we have plotted the density Eq. (3) for $Q = 3.22$ and $\sigma^2 = 0.85$: this is the theoretical value obtained assuming that the matrix is purely random except for its highest eigenvalue (dotted line). A better fit can be obtained with a smaller value of $\sigma^2 = 0.74$ (solid line), corresponding to 74% of the total variance. Inset: Same plot, but including the highest eigenvalue corresponding to the market, which is found to be 25 times greater than λ_{max}.

OCS

Which is the meaning of largest eigenvalues?

VASILIKI PLEROU *et al.*

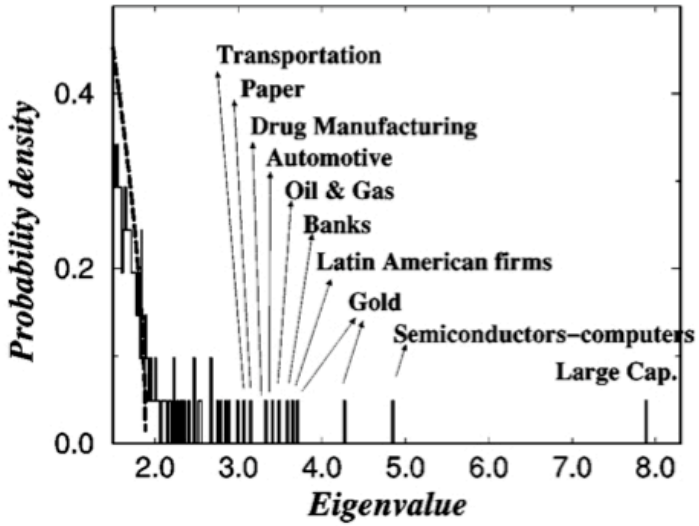


FIG. 13. Schematic illustration of the interpretation of the eigenvectors corresponding to the eigenvalues that deviate from the RMT upper bound. The dashed curve shows the RMT result of Eq. (6).

GOPIKRISHNAN, ROSENOW, PLEROU, AND STANLEY

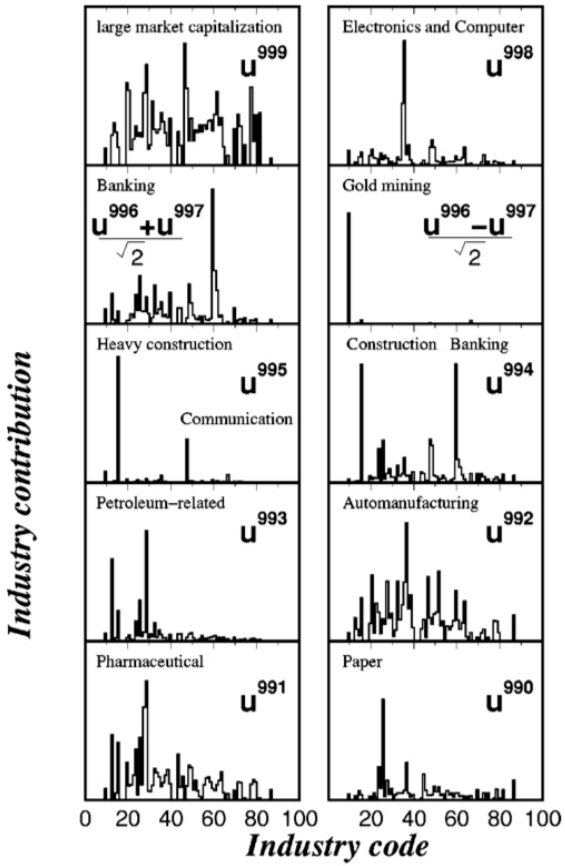


FIG. 1. Contribution X_l^k to industry sector l of eigenvector $\mathbf{u}^k$ for the deviating eigenvectors shows marked peaks at distinct values of SIC code, for all but $\mathbf{u}^{999}$ which contains stocks with large capitalizations as significant contributors.

Stanley and collaborators relate some of them to some economic sectors

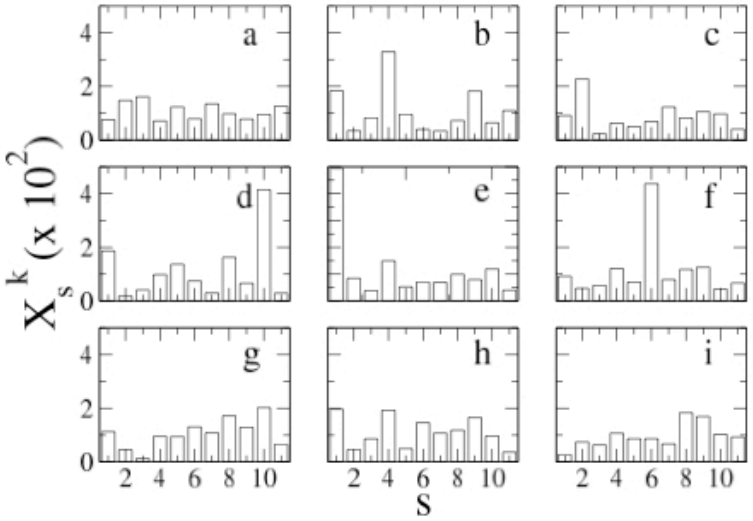
V- Plerou et al, Phys. Rev. E65, 066126 (2002)

P. Gopikrishnan et al, Phys. Rev. E64, 035106 (2001)

$$X_S^k = \sum_{i=1}^{n} P_{Si}\left[u_i^k\right]^2 \qquad P_{Si} = \begin{cases} \dfrac{1}{n_{Si}} \\ 0 \end{cases}$$

OCS

In a different market we[¶] were not able to reach the same conclusion. Only in a few cases components detected with PCA had a direct interpretation in terms of clear economic sectors.



f. Healthcare

	SECTOR	NUMBER
1	Technology	4
2	Financial	20
3	Energy	3
4	Consumer non-Cyclical	12
5	Consumer Cyclical	10
6	Healthcare	6
7	Basic Materials	5
8	Services	19
9	Utilities	6
10	Capital Goods	5
11	Transportation	2

Fig. 1. Contribution X_s^k of Eq. (3) for the first (a), second (b), third (c), sixth (f), seventh (g), eighth (h) and ninth (i) eigenvectors of the correlation matrix of daily returns of 92 LSE stocks. Panel (d) shows X_s^k for the linear combination $(\boldsymbol{u}^4 + \boldsymbol{u}^5)/\sqrt{2}$ and panel (e) for the linear combination $(\boldsymbol{u}^4 - \boldsymbol{u}^5)/\sqrt{2}$. The order of sectors is the same as in Table I.

Linear combination

d. Capital goods
e. Technology

[¶]C. Coronnello et al, Acta Phys. Pol. B36, 2653 (2005)

OCS

Random matrix theory as a filtering procedure

Filtered correlation matrix with RMT (1)

$\mathbf{C}$ = correlation matrix

$\mathbf{D}$ = diagonal matrix of eigenvalues of $\mathbf{C}$: $\mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_N]$

$\mathbf{V}$ = orthogonal matrix of eigenvectors of $\mathbf{C}$

Select λ_{max};

$$\mathbf{D}^* = \mathrm{diag}\left[\begin{cases} \lambda_i & \text{if } \lambda_i > \lambda_{max} \\ \lambda_0 & \text{otherwise} \end{cases}\right]$$

where λ_0 is the average of eigenvalues smaller than λ_{max}

$$\mathbf{C_B^t} = (c_{ij}^B) = \mathbf{V^T D^* V}$$

$$\mathbf{C_B} = \left(c_{ij}^B \Big/ \sqrt{c_{ii}^B c_{jj}^B} \right)$$

M. Potters, J.-P. Bouchaud & L. Laloux, Acta Phys. Pol. B 36 (9), pp. 2767-2784 (2005).

OCS

Filtered correlation matrix with RMT (2)

$\mathbf{C}$ = correlation matrix

$\mathbf{D}$ = diagonal matrix of eigenvalues of $\mathbf{C}$: $\mathrm{diag}[\lambda_1, \lambda_2, ..., \lambda_N]$

$\mathbf{V}$ = orthogonal matrix of eigenvectors of $\mathbf{C}$

Select λ_{max};

$$\mathbf{D}^* = \mathrm{diag}\left[\begin{cases} \lambda_i & \text{if } \lambda_i > \lambda_{max} \\ 0 & \text{otherwise} \end{cases}\right]$$

$$\mathbf{C}_S^t = (c_{ij}^S) = \mathbf{V}^T \mathbf{D}^* \mathbf{V}$$

$$\mathbf{C}_S = \left(\delta_{ij} + c_{ij}^S \left[1 - \delta_{ij}\right]\right)$$

B. Rosenow, V. Plerou, P. Gopikrishnan & H.E. Stanley, Europhys. Lett. 59 (4), pp. 500-506 (2002)

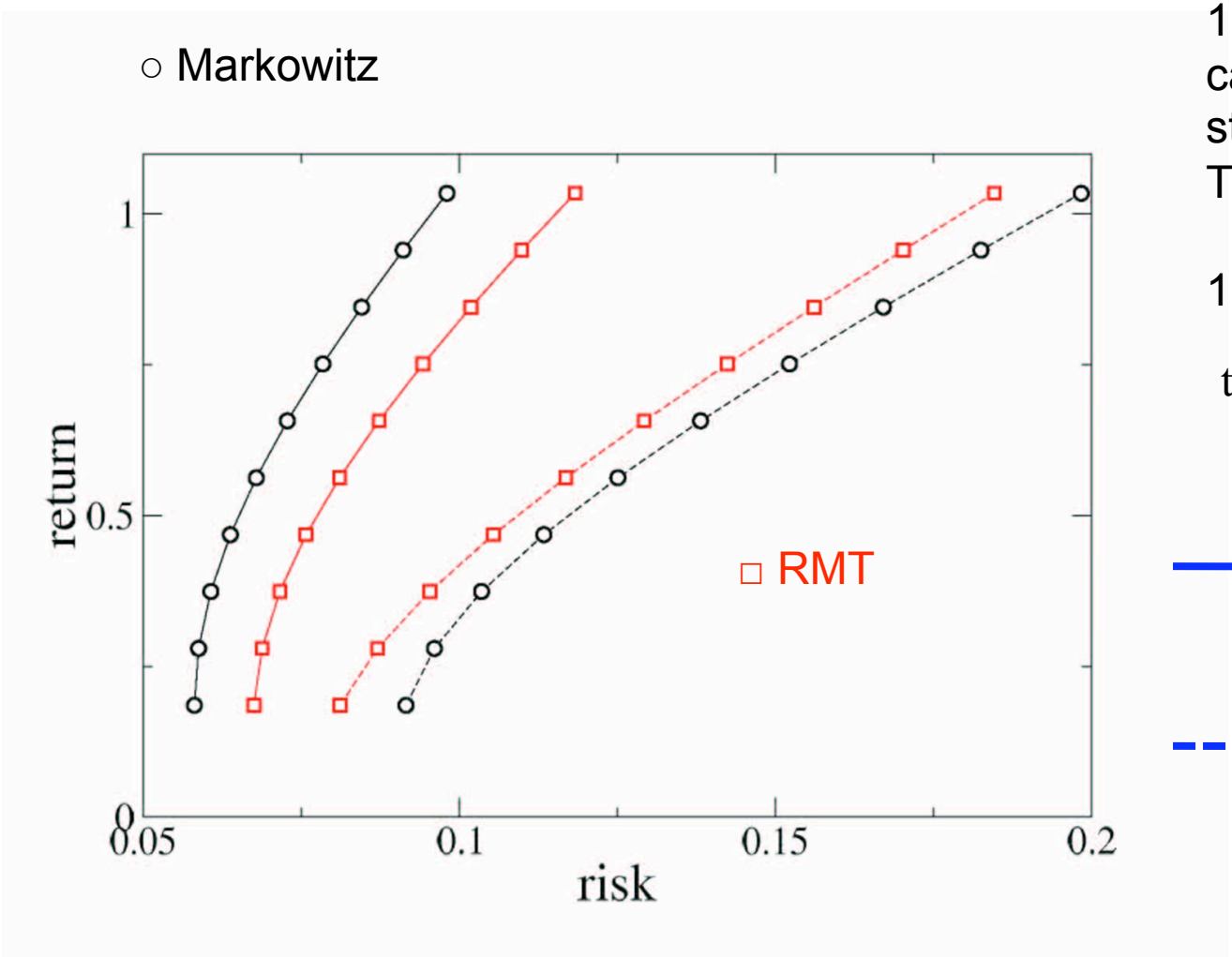
OCS

Optimal portfolios with RMT

B. Rosenow et al, Europhys. Lett. **59**, 500 (2002) have shown that the distance between the optimal predicted return-risk profile and the realized one is (often) larger in the Markowitz case than in a case when the correlation coefficient matrix is filtered by using the results of the RMT.

OCS

A (good) result



150 most capitalized stocks
T=500

1989-1992

t_o=Jan 1990

——— expected $(t_o\text{-}500, t_o)$

- - - - - realized (t_o, t_o+500)

OCS

Another filtering procedure: Hierarchical clustering

By starting from a correlation matrix (which is a similarity measure)

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.413	0.518	0.543	0.529	0.341	0.271	0.231	0.412	0.294
IBM		1	0.471	0.537	0.617	0.552	0.298	0.475	0.373	0.270
BAC			1	0.547	0.591	0.400	0.258	0.349	0.370	0.276
AXP				1	0.664	0.422	0.347	0.351	0.414	0.269
MER					1	0.533	0.344	0.462	0.440	0.318
TXN						1	0.305	0.582	0.355	0.245
SLB							1	0.193	0.533	0.592
MOT								1	0.258	0.166
RD									1	0.590
OXY										1

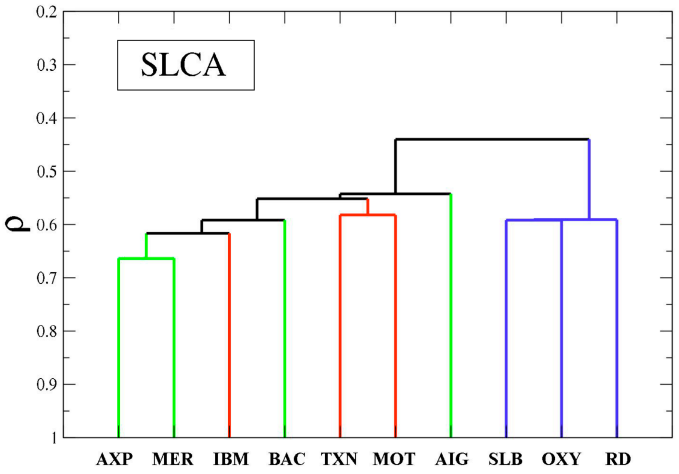
AXP	MER	0.664
IBM	MER	0.617
SLB	OXY	0.592
BAC	MER	0.591
RD	OXY	0.590
TXN	MOT	0.582
IBM	TXN	0.552
AXP	BAC	0.547
AIG	AXP	0.543
AXP	IBM	0.537
SLB	RD	0.533
MER	TXN	0.533
AIG	MER	0.529
AIG	BAC	0.518
IBM	MOT	0.475
MOT	MER	0.462
MER	RD	0.440
AXP	TXN	0.422
.....		

OCS

Single linkage

One may obtain a simplified matrix by using classical clustering methods such us the single linkage clustering

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.543	0.543	0.543	0.543	0.543	0.440	0.543	0.440	0.440
IBM		1	0.591	0.617	0.617	0.552	0.440	0.552	0.440	0.440
BAC			1	0.591	0.591	0.552	0.440	0.552	0.440	0.440
AXP				1	0.664	0.552	0.440	0.552	0.440	0.440
MER					1	0.552	0.440	0.552	0.440	0.440
TXN						1	0.440	0.582	0.440	0.440
SLB							1	0.440	0.590	0.592
MOT								1	0.440	0.440
RD									1	0.590
OXY										1



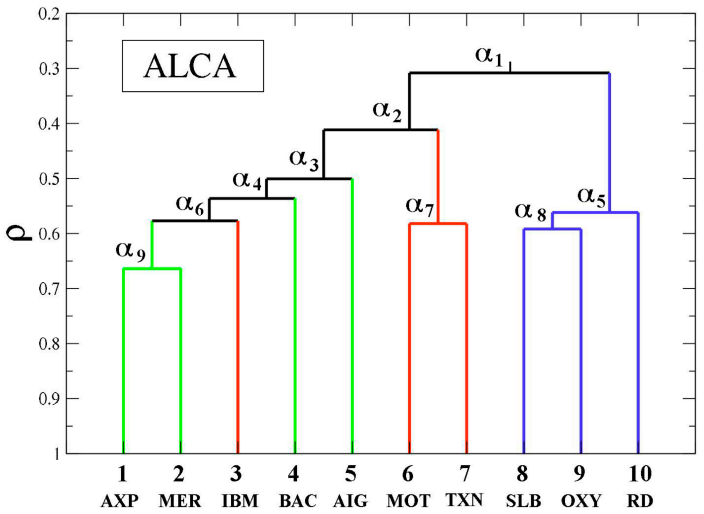
$C^{<}_{SL}$

OCS

Average linkage

Or, for example, the average linkage clustering

	AIG	IBM	BAC	AXP	MER	TXN	SLB	MOT	RD	OXY
AIG	1	0.501	0.501	0.501	0.501	0.412	0.308	0.412	0.308	0.308
IBM		1	0.536	0.577	0.577	0.412	0.308	0.412	0.308	0.308
BAC			1	0.536	0.536	0.412	0.308	0.412	0.308	0.308
AXP				1	0.664	0.412	0.308	0.412	0.308	0.308
MER					1	0.412	0.308	0.412	0.308	0.308
TXN						1	0.308	0.582	0.308	0.308
SLB							1	0.308	0.562	0.591
MOT								1	0.308	0.308
RD									1	0.562
OXY										1

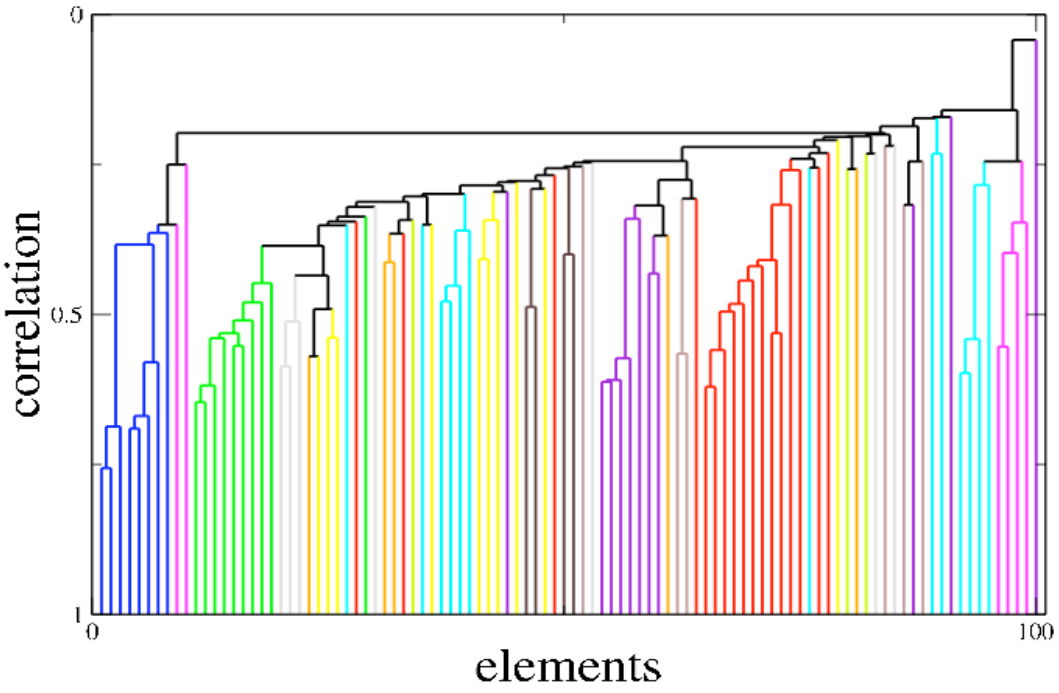


$C^{<}_{AL}$

OCS

Hierarchical clustering output in a typical case

$N = 100$ (NYSE) daily returns 1995 - 1998



Average Linkage Cluster Analysis

$$C^{<} = (\rho_{ij}^{<})$$

$$\rho_{ij}^{<} = \rho_{\alpha_k}$$

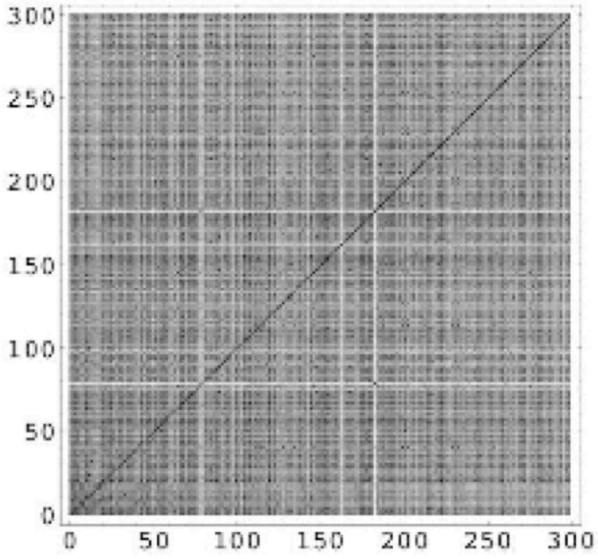
where

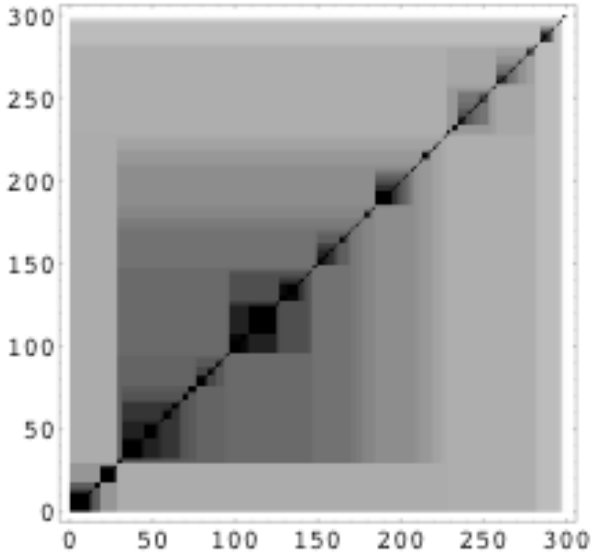
α_k

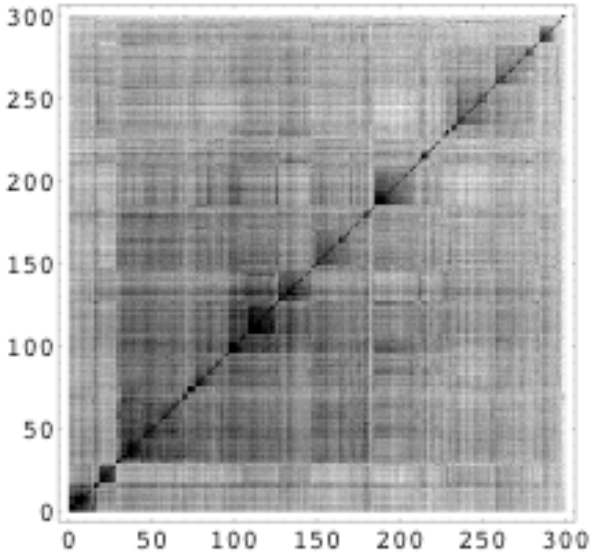
is the first node where elements *i* and *j* merge together

Filtered matrix

$N = 300$ (NYSE); daily returns 2001 - 2003

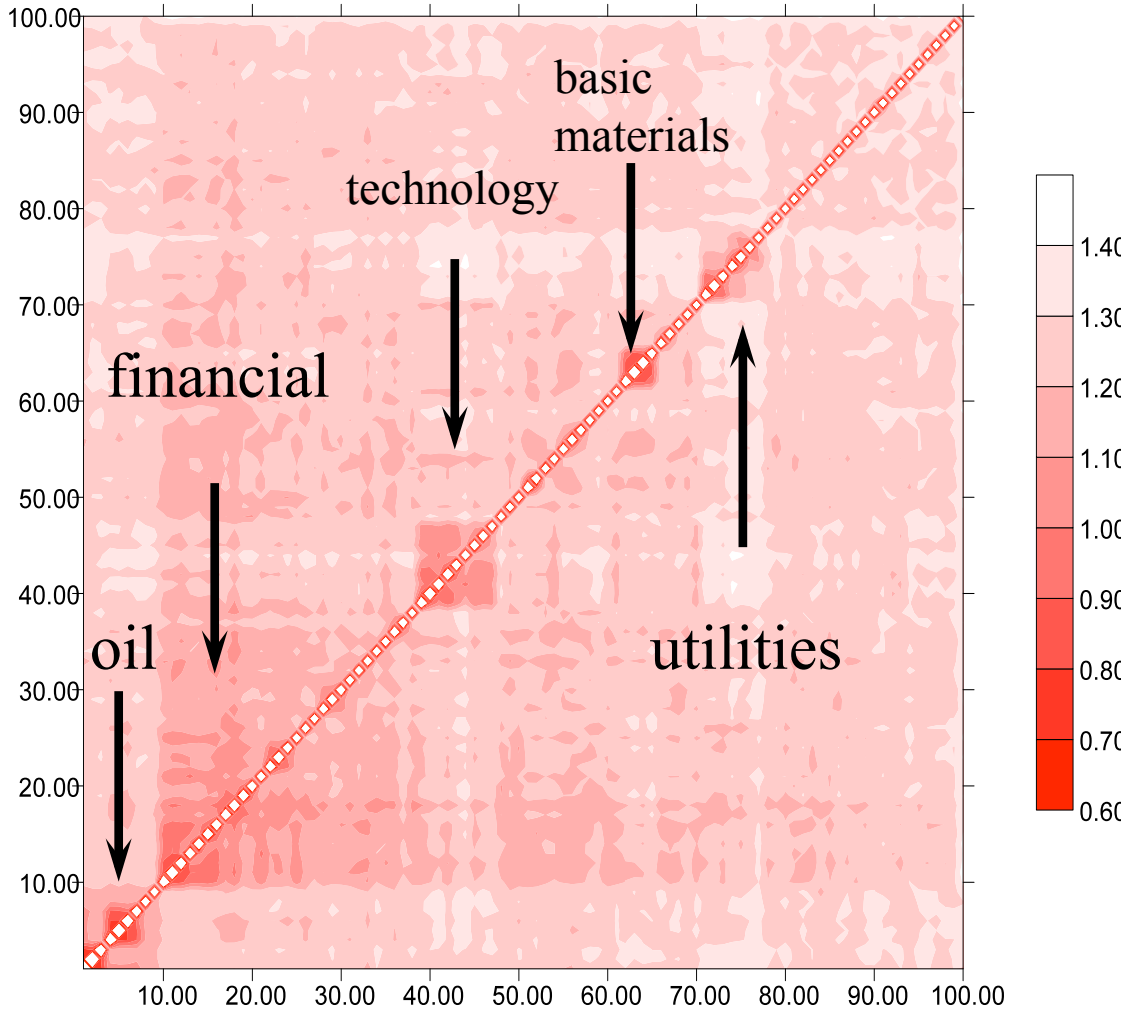






C $C^{<}$ from ALCA $C_{ht \text{ ordered}}$

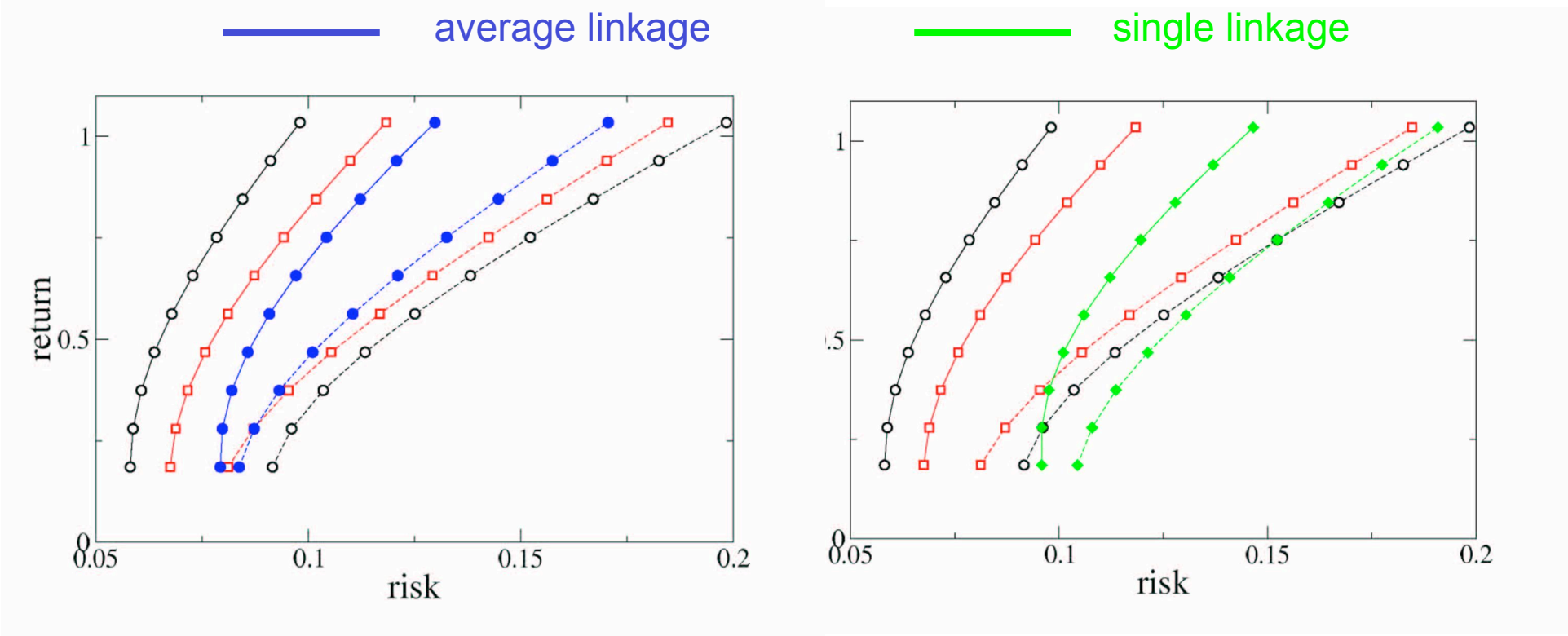
The complete matrix is richer of information



When one uses the stock order of the hierarchical tree the correlation matrix assumes a better readability

OCS

Filtering by cluster analysis: Portfolio optimization



150 most capitalized stocks T=500 1989-1992

V. Tola, F. Lillo, M. Gallegati, R.N.M., Cluster analysis for portfolio optimization, Journal of Economic Dynamics & Control, 32, 235-258 (2008)

OCS

The shrinkage procedure

$$\mathbf{C}(\alpha) = \alpha\mathbf{T} + (1-\alpha)\mathbf{S}$$

S = sample correlation matrix

T = target matrix

(typically very stable to statistical fluctuations)

T could be the identity matrix, the constant correlation matrix, etc.

α is determined by expectation/optimization procedures

O. Ledoit, M. Wolf, J. Mult. Analysis **88**, 365 (2004).

OCS

Some values of α optimal

Target A: "diagonal, unit variance"
0 estimated parameters

$$t_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$\hat{\lambda}^\star = \frac{\sum_{i \neq j} \widehat{\mathrm{Var}}(s_{ij}) + \sum_i \widehat{\mathrm{Var}}(s_{ii})}{\sum_{i \neq j} s_{ij}^2 + \sum_i (s_{ii} - 1)^2}$$

Target B: "diagonal, common variance"
1 estimated parameter: v

$$t_{ij} = \begin{cases} v = \mathrm{avg}(s_{ii}) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$\hat{\lambda}^\star = \frac{\sum_{i \neq j} \widehat{\mathrm{Var}}(s_{ij}) + \sum_i \widehat{\mathrm{Var}}(s_{ii})}{\sum_{i \neq j} s_{ij}^2 + \sum_i (s_{ii} - v)^2}$$

Target C: "common (co)variance"
2 estimated parameters: v, c

$$t_{ij} = \begin{cases} v = \mathrm{avg}(s_{ii}) & \text{if } i = j \\ c = \mathrm{avg}(s_{ij}) & \text{if } i \neq j \end{cases}$$

$$\hat{\lambda}^\star = \frac{\sum_{i \neq j} \widehat{\mathrm{Var}}(s_{ij}) + \sum_i \widehat{\mathrm{Var}}(s_{ii})}{\sum_{i \neq j} (s_{ij} - c)^2 + \sum_i (s_{ii} - v)^2}$$

Target D: "diagonal, unequal variance"
p estimated parameters: s_{ii}

$$t_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$\hat{\lambda}^\star = \frac{\sum_{i \neq j} \widehat{\mathrm{Var}}(s_{ij})}{\sum_{i \neq j} s_{ij}^2}$$

Target E: "perfect positive correlation"
p estimated parameters: s_{ii}

$$t_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ \sqrt{s_{ii} s_{jj}} & \text{if } i \neq j \end{cases}$$

$$f_{ij} = \tfrac{1}{2}\{ \sqrt{\tfrac{s_{jj}}{s_{ii}}} \widehat{\mathrm{Cov}}(s_{ii}, s_{ij}) + \sqrt{\tfrac{s_{ii}}{s_{jj}}} \widehat{\mathrm{Cov}}(s_{jj}, s_{ij}) \}$$

$$\hat{\lambda}^\star = \frac{\sum_{i \neq j} \widehat{\mathrm{Var}}(s_{ij}) - f_{ij}}{\sum_{i \neq j} (s_{ij} - \sqrt{s_{ii} s_{jj}})^2}$$

Target F: "constant correlation"
$p + 1$ estimated parameters: $s_{ii}, \bar{r}$

$$t_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ \bar{r} \sqrt{s_{ii} s_{jj}} & \text{if } i \neq j \end{cases}$$

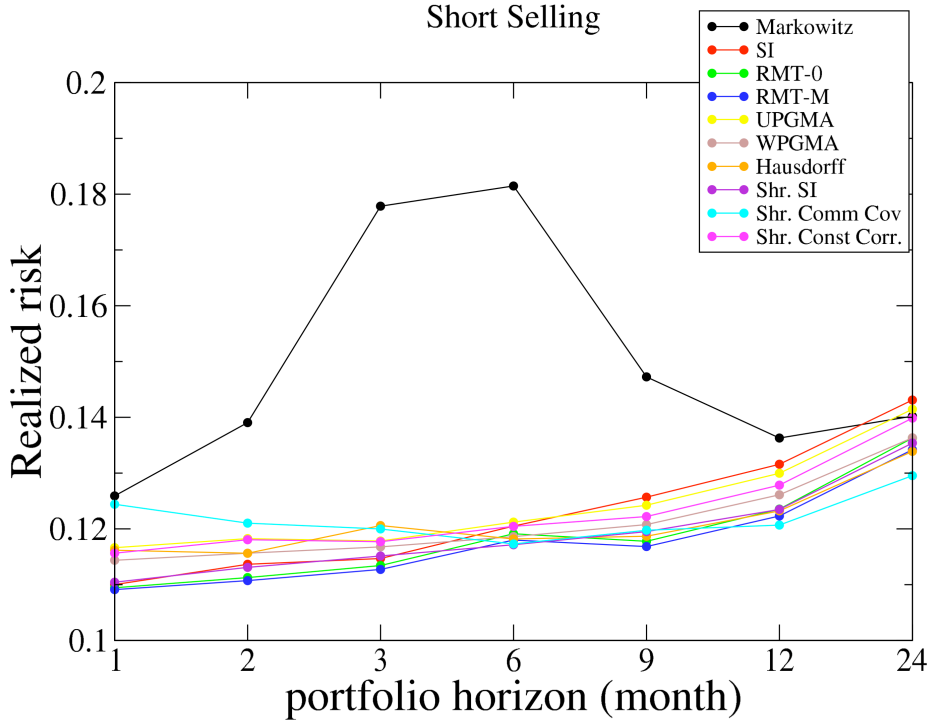
$$\hat{\lambda}^\star = \frac{\sum_{i \neq j} \widehat{\mathrm{Var}}(s_{ij}) - \bar{r} f_{ij}}{\sum_{i \neq j} (s_{ij} - \bar{r} \sqrt{s_{ii} s_{jj}})^2}$$

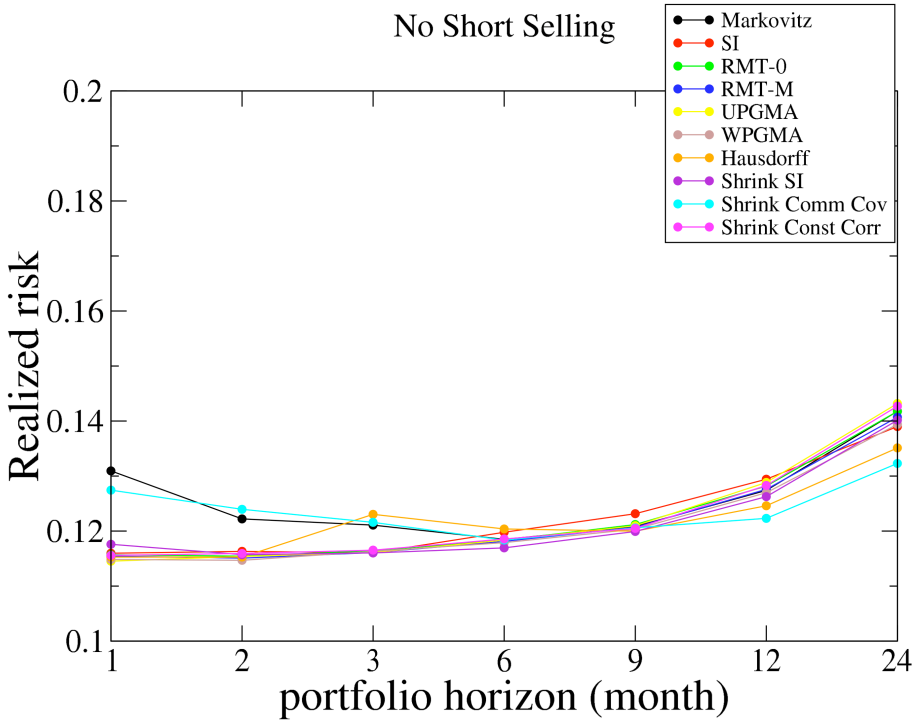
Table 2: Six commonly used shrinkage targets for the covariance matrix and associated estimators of the optimal shrinkage intensity – see main text for discussion. *Abbreviations:* v, average of sample variances; c, average of sample covariances; $\bar{r}$, average of sample correlations.

J. Schäfer, K. Strimmer, Stat. Appl. Gen. Mol. Biol. 4, 32 (2005).).

Comparison of several methods (realized risk)

N=90 Stocks (US equity markets, blue chips)





Markowitz Single Index RMT-0 RMT-M UPGMA WPGMA

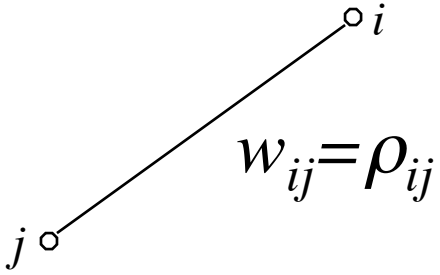
Hausdorff Shr. Sin. Ind. Shr. Comm. Cov. Shr. Comm. Corr.

E. Pantaleo, M. Tumminello, F. Lillo, RNM, When do improved covariance matrix estimators enhance portfolio optimization? An empirical comparative study of nine estimators, Quantitative Finance 11, 1067-1080 (2011)

OCS

Correlation based networks

$$(i, j, \rho_{ij})$$



$$C = \begin{pmatrix} 1 & 0.13 & 0.90 & 0.81 \\ 0.13 & 1 & 0.57 & 0.34 \\ 0.90 & 0.57 & 1 & 0.71 \\ 0.81 & 0.34 & 0.71 & 1 \end{pmatrix} \rightarrow S = \begin{pmatrix} 1 & 3 & 0.90 \\ 1 & 4 & 081 \\ 3 & 4 & 0.71 \\ 2 & 3 & 0.57 \\ 2 & 4 & 0.34 \\ 1 & 2 & 0.13 \end{pmatrix}$$

Correlation Matrix (C)

Sorted List of Links (S)

OCS

Minimum Spanning Tree

Define a **similarity measure between the elements of the system**

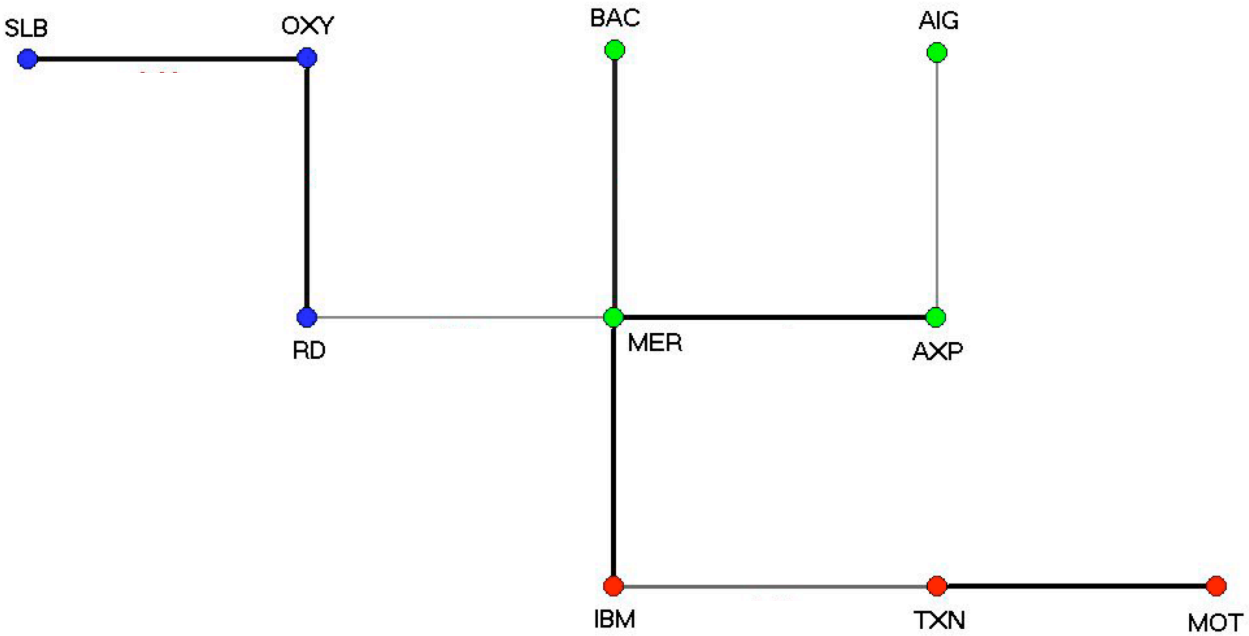
Construct the list ***S* by ordering similarities in decreasing order**

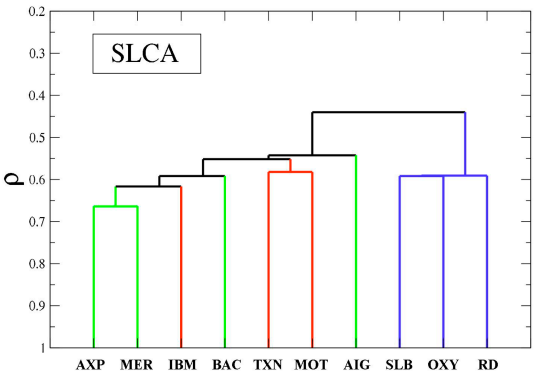
Starting from the first
element of *S*,
add the corresponding **link**
if and only if
the graph is still a Forest or a Tree

Minimum Spannig Tree
MST

Correlation based tree(s)

For the single linkage clustering procedure the correlation based tree is the minimum spanning tree

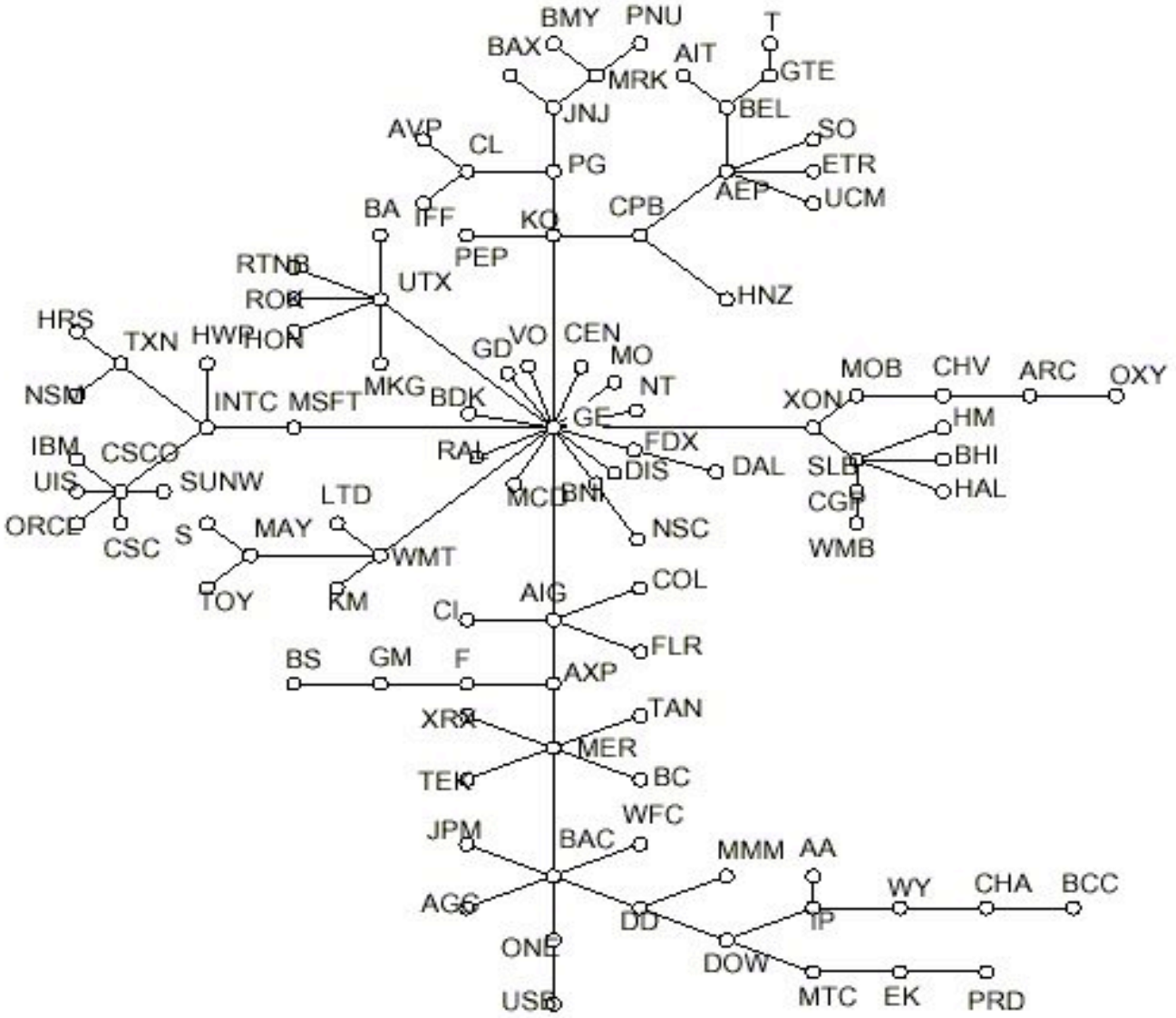




Correlation based trees and hierarchical trees do NOT carry the same amount of information.

OCS

Minimum Spanning Tree (MST)



N=100 (NYSE)
daily returns
1995-1998
T=1011

R. N. Mantegna, EPJ B 11, 193 (1999) G. Bonanno, F. Lillo and R.N.M., Quant. Fin. 1, 96 (2001)

MST and Planar Maximally Filtered Graph (PMFG)

Define a **similarity measure between the elements of the system**

Construct the list ***S* by ordering similarities in decreasing order**

Starting from the first element of *S*, **add** the corresponding **link** **if** and only if **the graph is still a Forest or a Tree**

Minimum Spannig Tree
MST

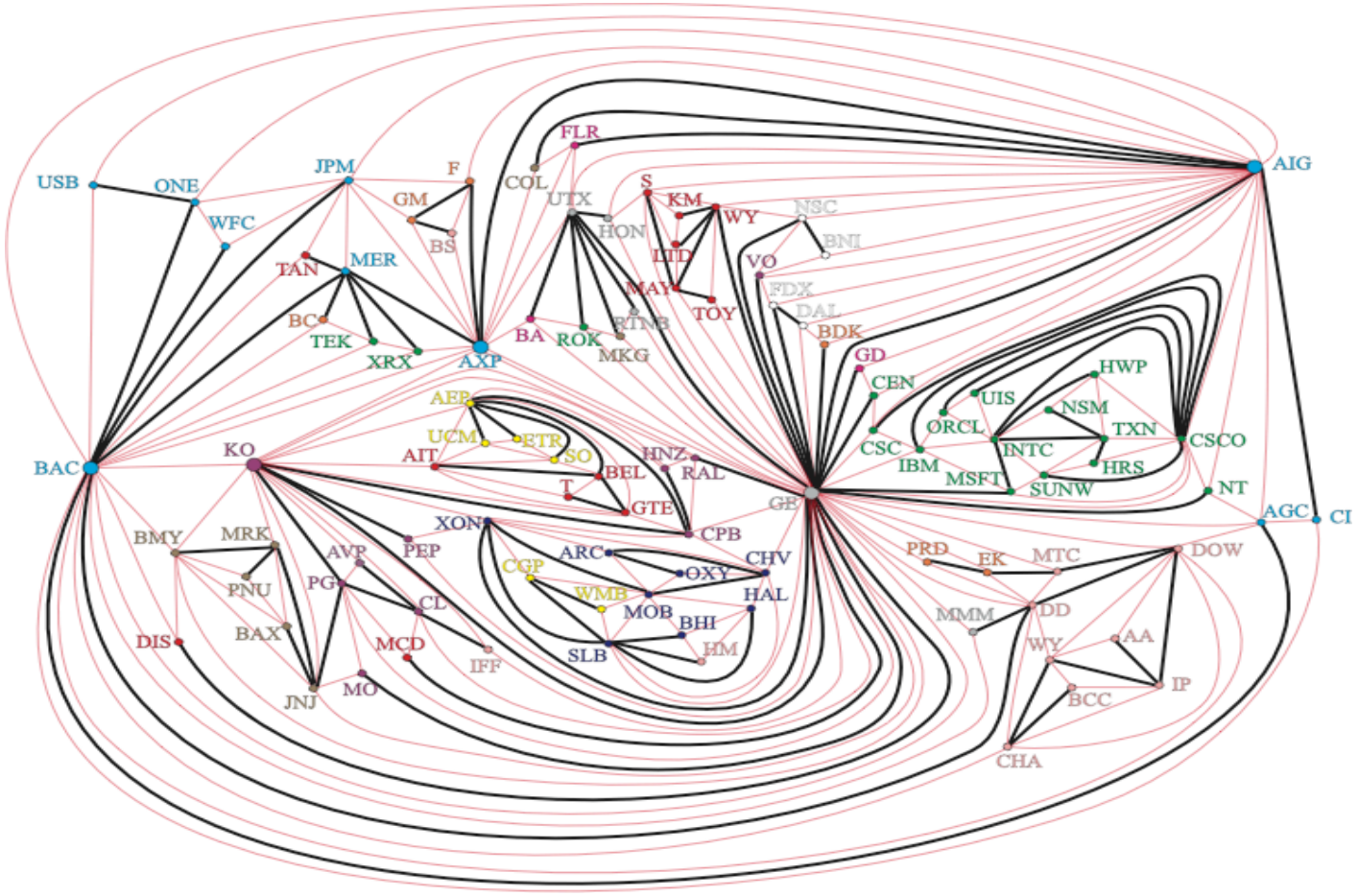
Starting from the first element of *S*, **add** the corresponding **link** **if** and only if **the graph is still Planar (g=0)**

Planar Maximally Filtered Graph
PMFG

M. Tumminello, T. Di Matteo, T. Aste and R.N.M., PNAS USA 102, 10421 (2005)

OCS

Planar Maximally Filtered Graph

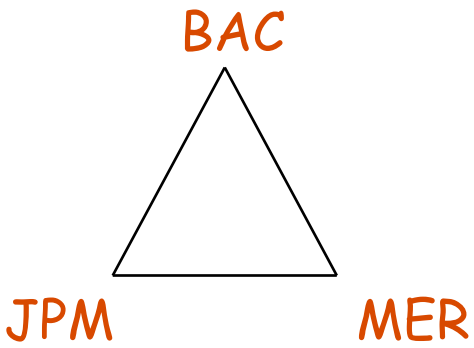


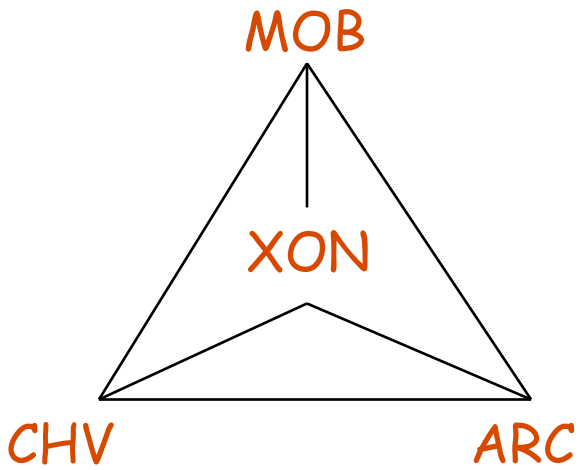
N=100 (NYSE)
daily returns
1995-1998
T=1011

M. Tumminello, T. Di Matteo, T. Aste and R.N.M., PNAS USA 102, 10421 (2005)

Cliques in the PMFG

When g=0, topological constraints allow the observation of **cliques** of 3 and 4 vertices only.





OCS

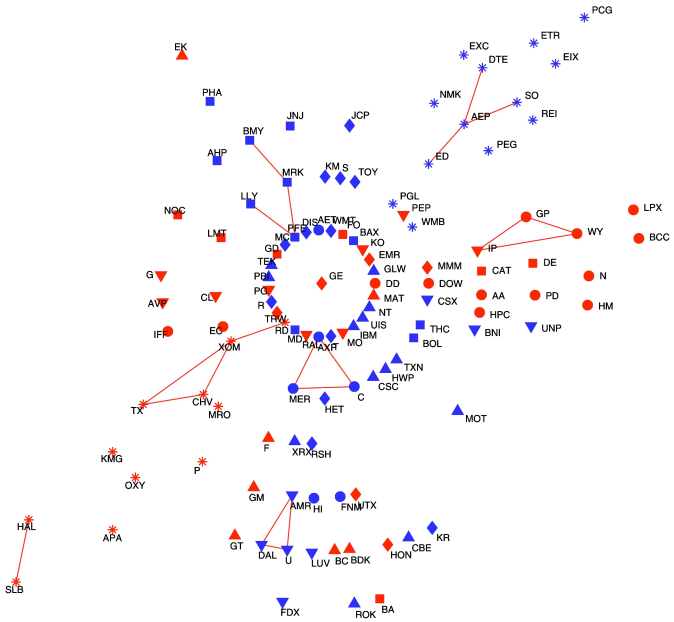


Fig. 1. Sample graph of $N = 116$ vertices and $n = 20$ edges, corresponding to a connection probability $p = n/[N(N-1)/2] \approx 0.003$.

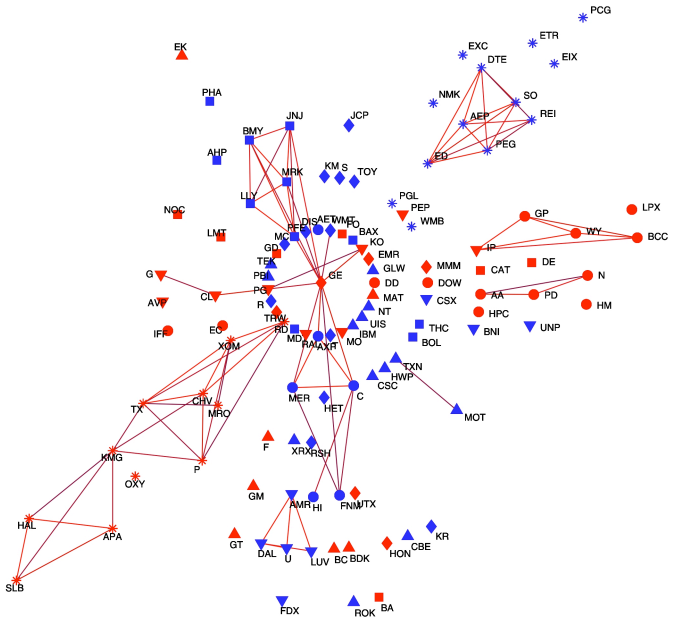


Fig. 3. Sample graph for $n = 80$ edges ($p \approx 0.012$).

Other correlation based networks have been obtained by selecting a thresholds on the similarity measure.

J.-P. Onnela, K. Kaski, and J. Kertész, Eur. Phys. J. B 38, 353–362 (2004)

OCS

Network analysis of a single index model: Topology of MSTs

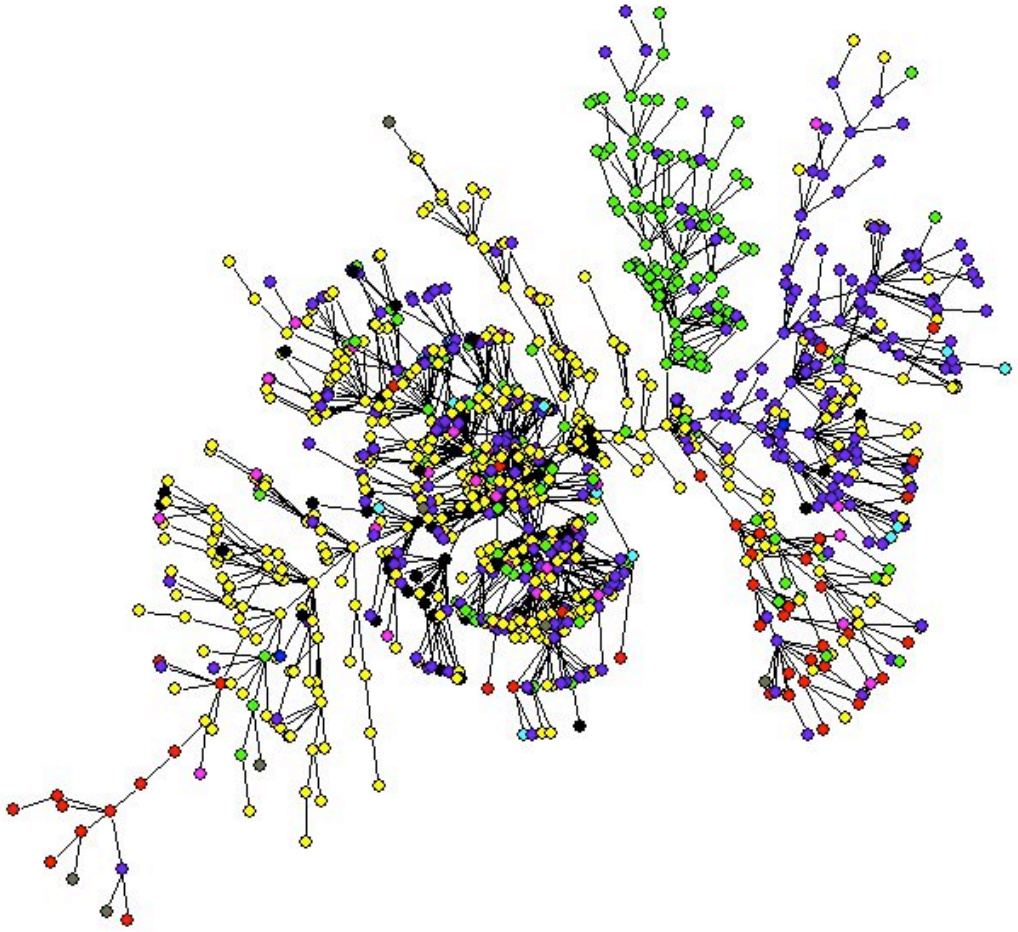
Topology† of MSTs in

- empirical data;
- in a one-factor model

†Bonanno, Caldarelli, Lillo and Mantegna, PRE **68**, 046130 (2003).

OCS

Empirical data



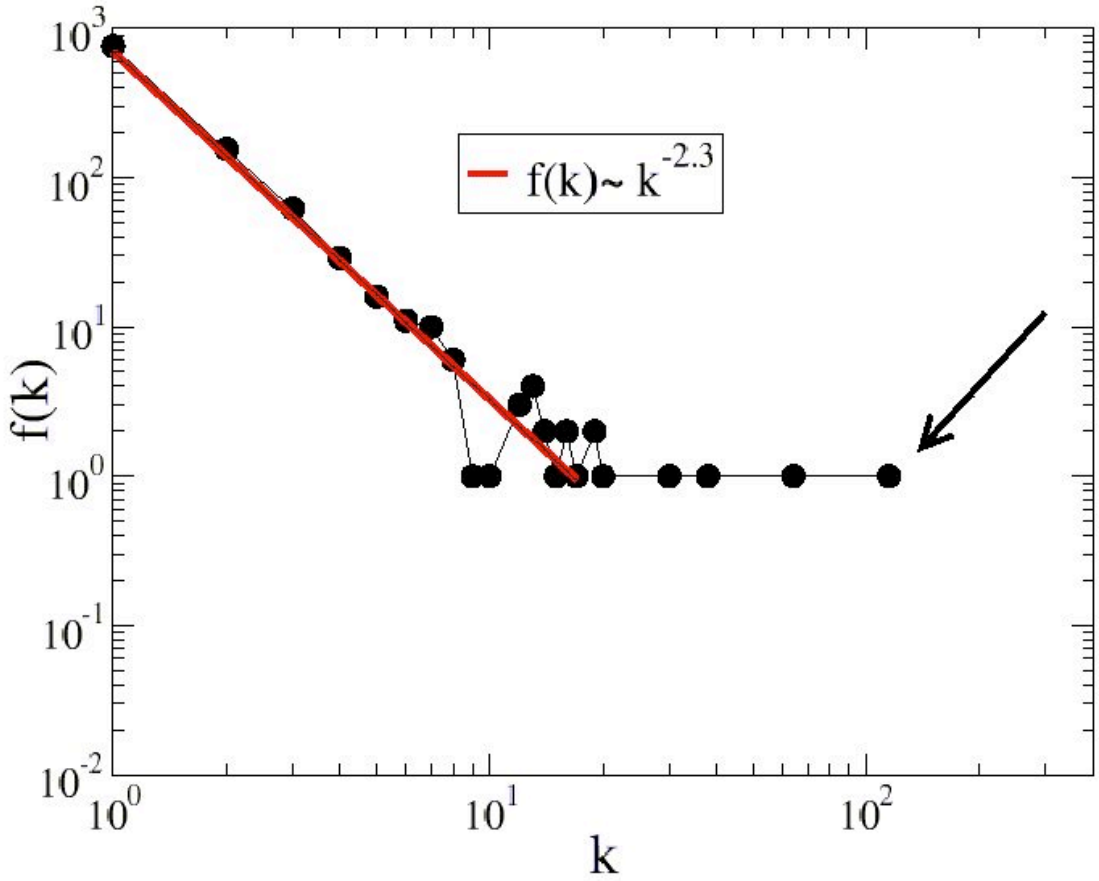
1071 stocks continuously traded at the NYSE during the period 1987-1998 (3030 trading days)

The color code refers to the SIC index:

- finance
- manufacturing
- construction
- utilities
- wholesale trade
- mining
- retail trade
- services
- public administration

OCS

Empirical data: degree distribution



It is *approximately* power-law decaying for low k values but a few nodes with very high degree are present.

OCS

Comparison with the one-factor model

We compare our empirical results with numerical simulations based on the **one-factor model**

$$R_i(t) = \alpha_i + \beta_{iM} R_M(t) + \varepsilon_i(t)$$

where α_i and β_{iM} are two real parameters, ε_i is a zero-mean noise term characterized by a variance equal to σ_i^2

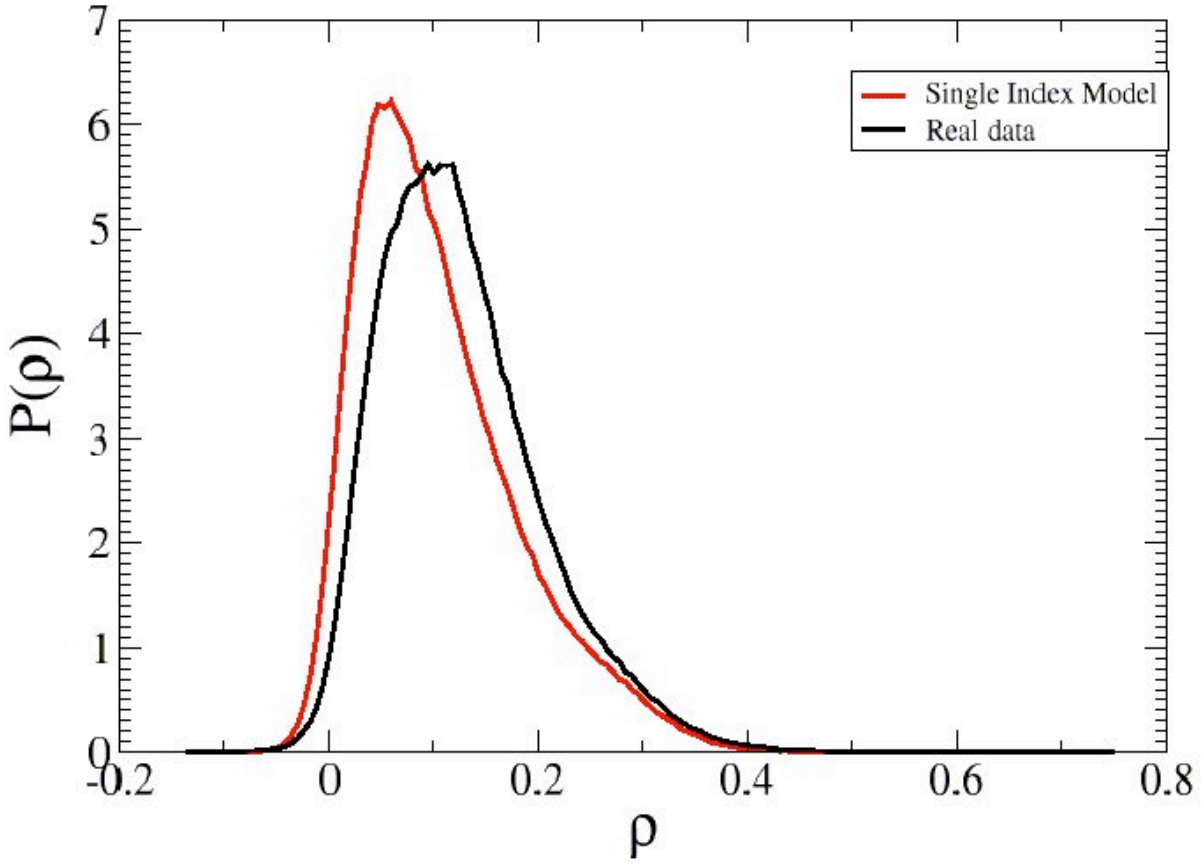
$R_M(t)$ is the market factor. We choose it as the SP 500 index.

In our investigation, we estimate the model parameters and generate an artificial market.

OCS

Correlation in the one-factor model

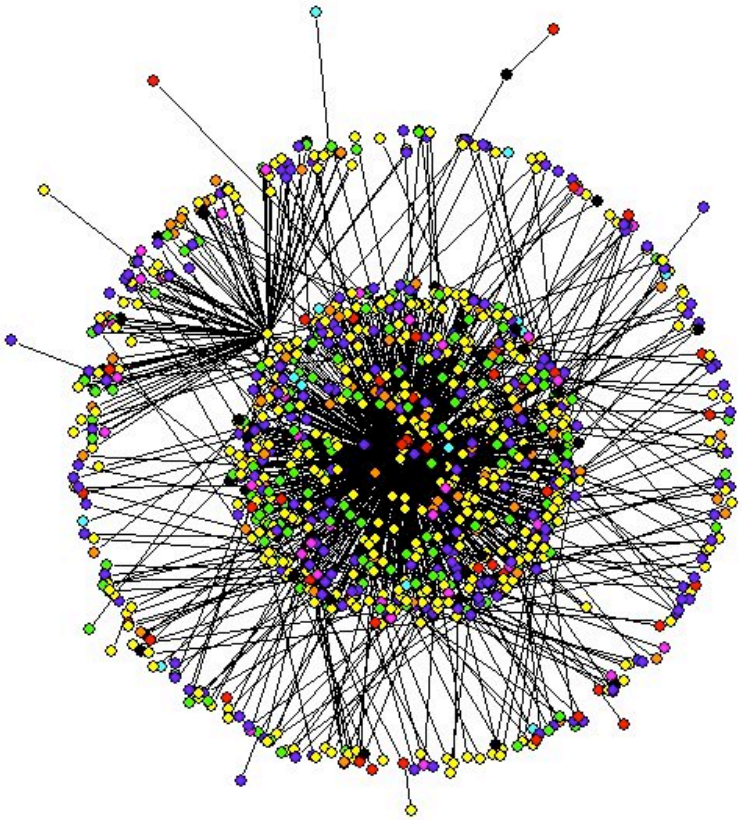
The one-factor model explains more than 85% of the elements of correlation matrix



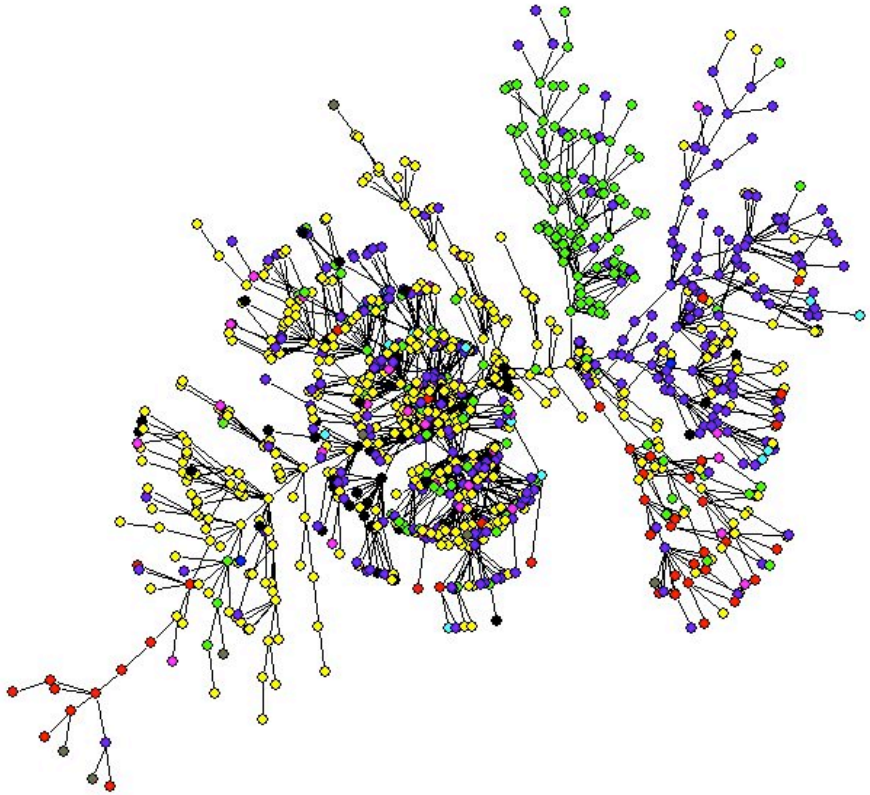
OCS

MST of a one-factor model

One-factor model



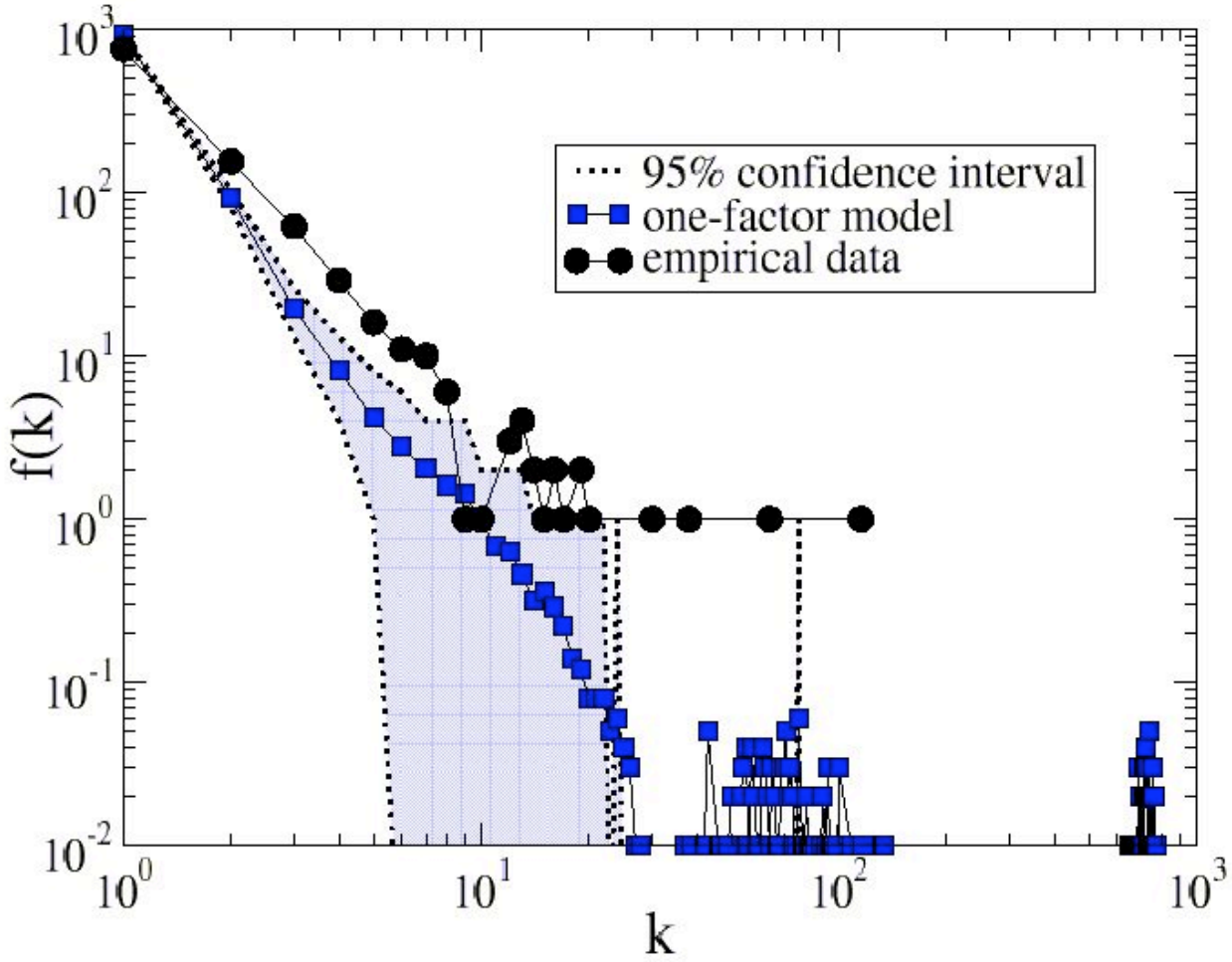
Real data



OCS

Topological properties of the one-factor model

Degree distribution



OCS

How to assess the stability of the information filtered out by the hierarchical network and correlation based network?

OCS

A validation based on bootstrap

Data Set

	e_1	e_2	e_3	…	e_n
t_1	0.113	1.123	-0.002	…	0.198
t_2	1.567	0.789	0.842	…	-0.234
t_3	1.065	-1.962	0.567	…	1.785
t_4	1.112	0.998	-0.424	…	2.735
t_5	-0.211	0.312	-0217	…	0.587
…	…	…	…	…	…
T	0.479	-1.828	-2.041	…	-0.193

Pseudo-replicate Data Set

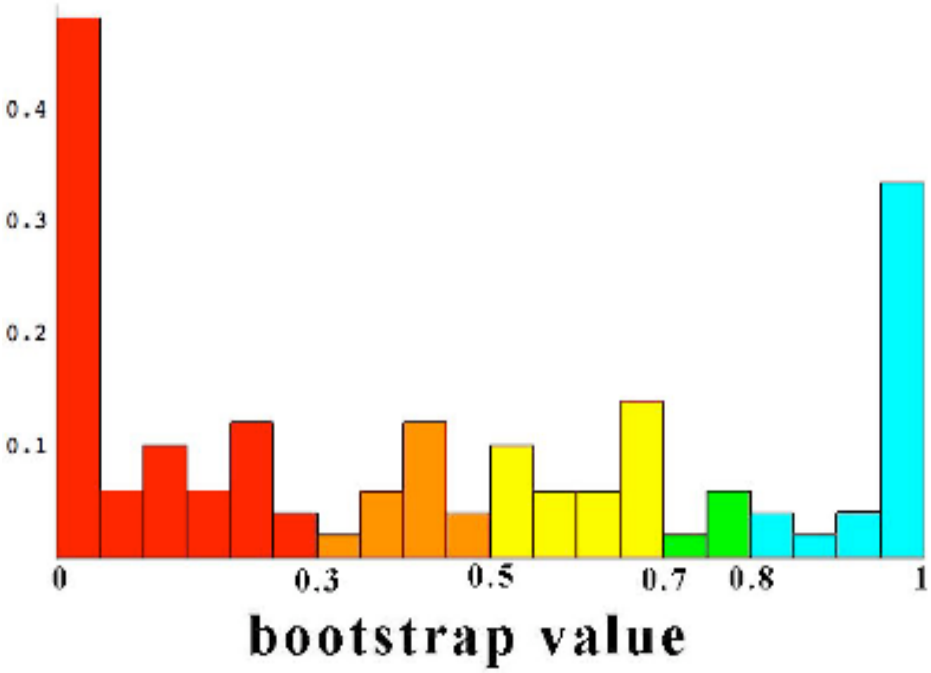
e_1	e_2	e_3	…	e_n
1.567	0.789	0.842	…	-0.234
0.113	1.123	-0.002	…	0.198
1.065	-1.962	0.567	…	1.785
0.113	1.123	-0.002	…	0.198
0.479	-1.828	-2.041	…	-0.193
…	…	…	…	…
0.479	-1.828	-2.041	…	-0.193

M surrogated data matrices are constructed, e.g. M=1000.

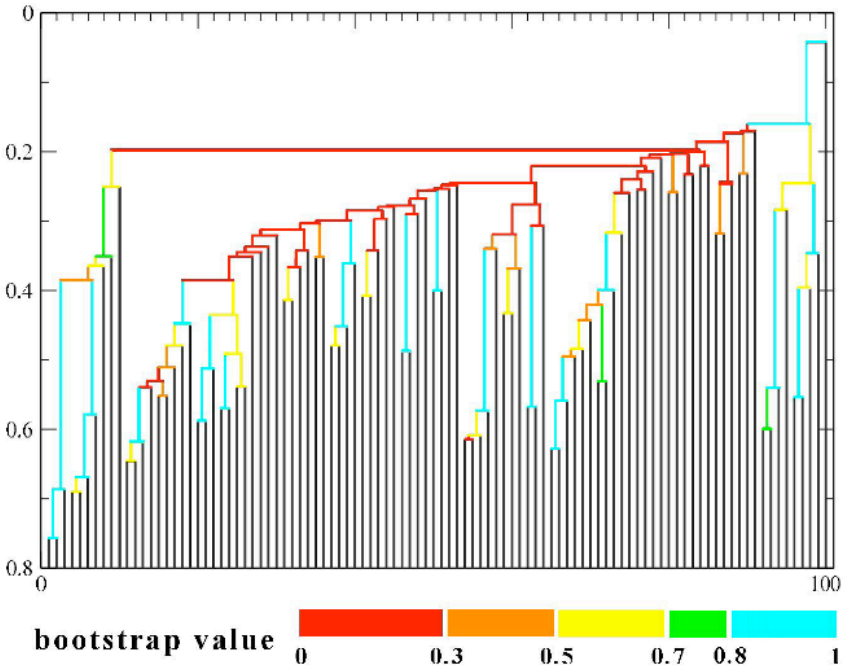
OCS

Bootstrap value of nodes of hierarchical trees

Bootstrap value density



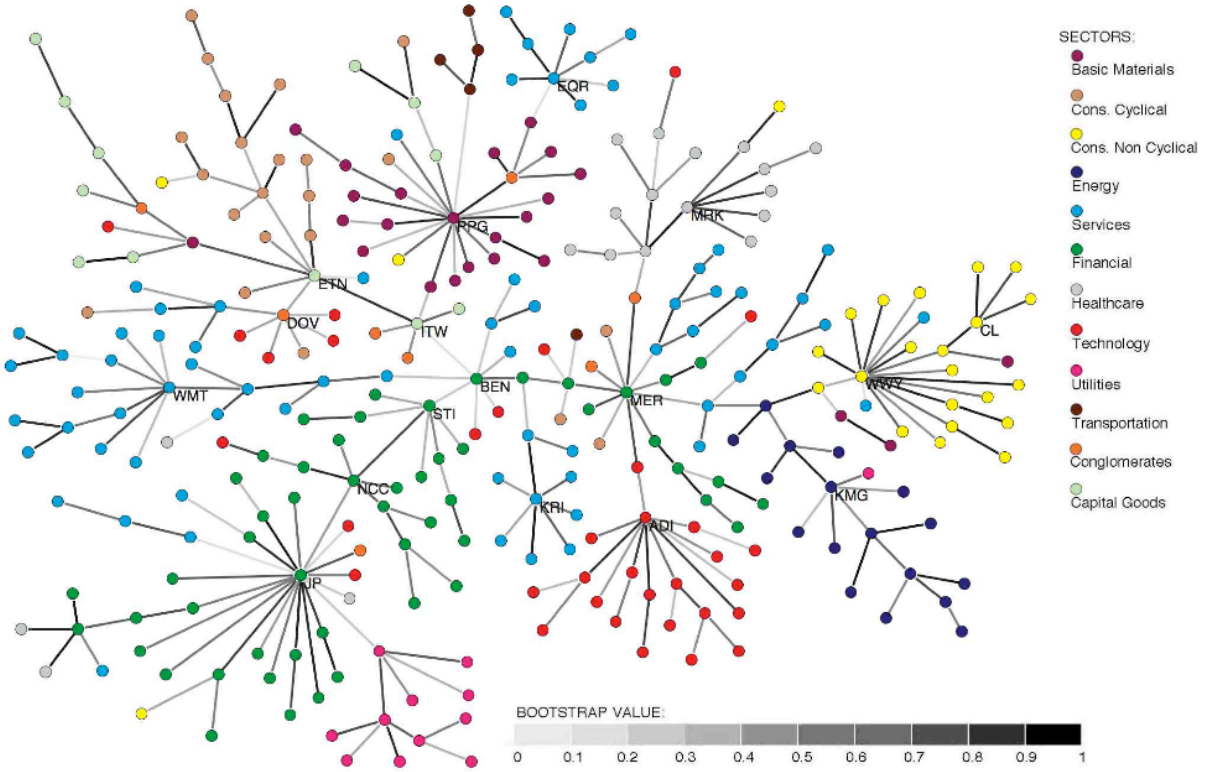
ALCA



M. Tumminello, C. Coronnello, S. Micciché, F. Lillo and R.N.M., Int. J. Bifurcation Chaos **17,** 2319-2329 (2007).

OCS

Statistical reliability of the minimum spanning tree



$N = 300$ (NYSE)

daily returns

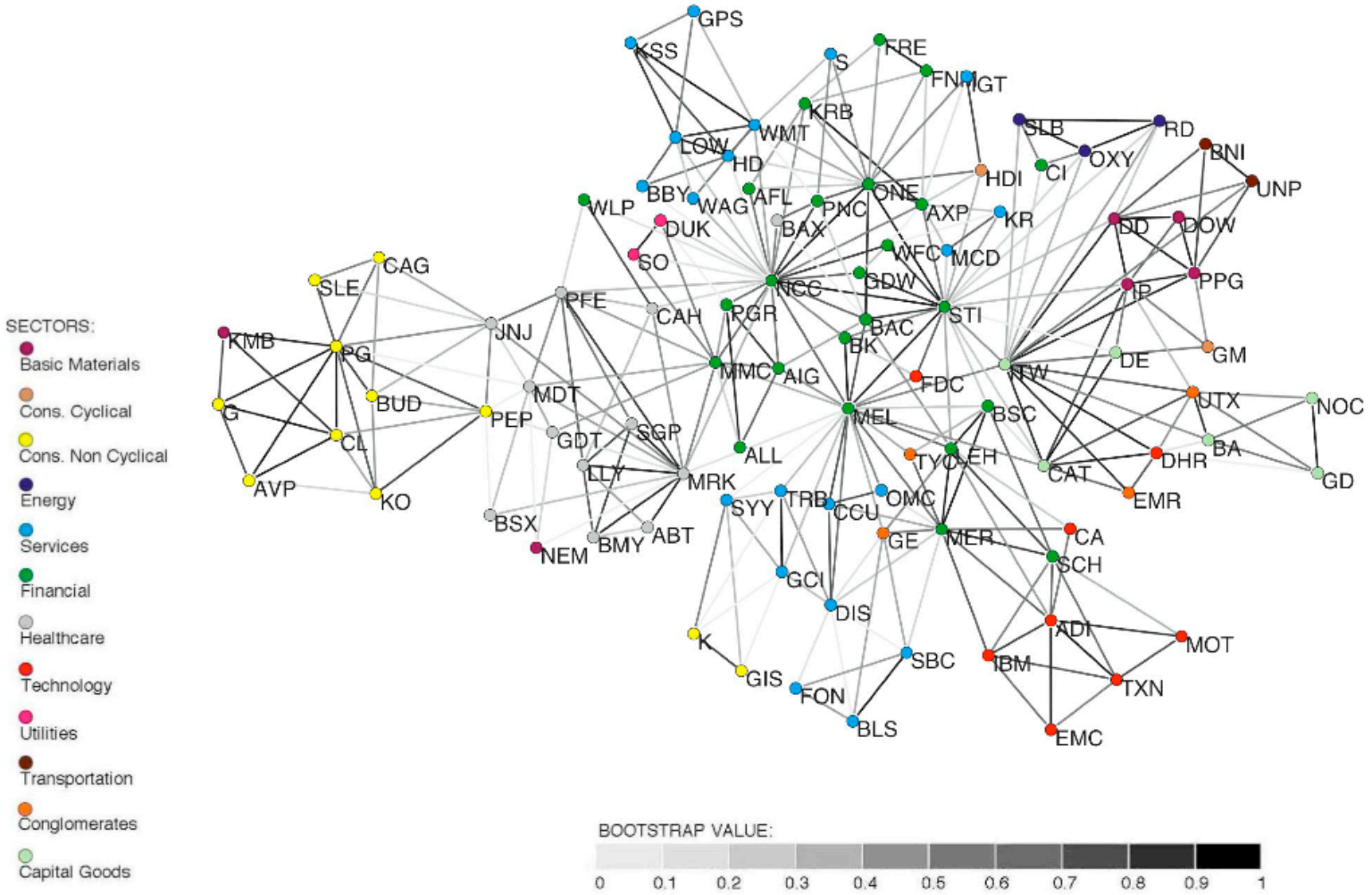
2001 - 2003

$T = 748$

M. Tumminello, C. Coronnello, S. Micciché, F. Lillo and R.N.M., Int. J. Bifurcation Chaos **17,** 2319-2329 (2007).

OCS

Statistical reliability of the planar maximally filtered graph



NYSE

$N = 100$

daily returns

2002

$T = 256$

OCS

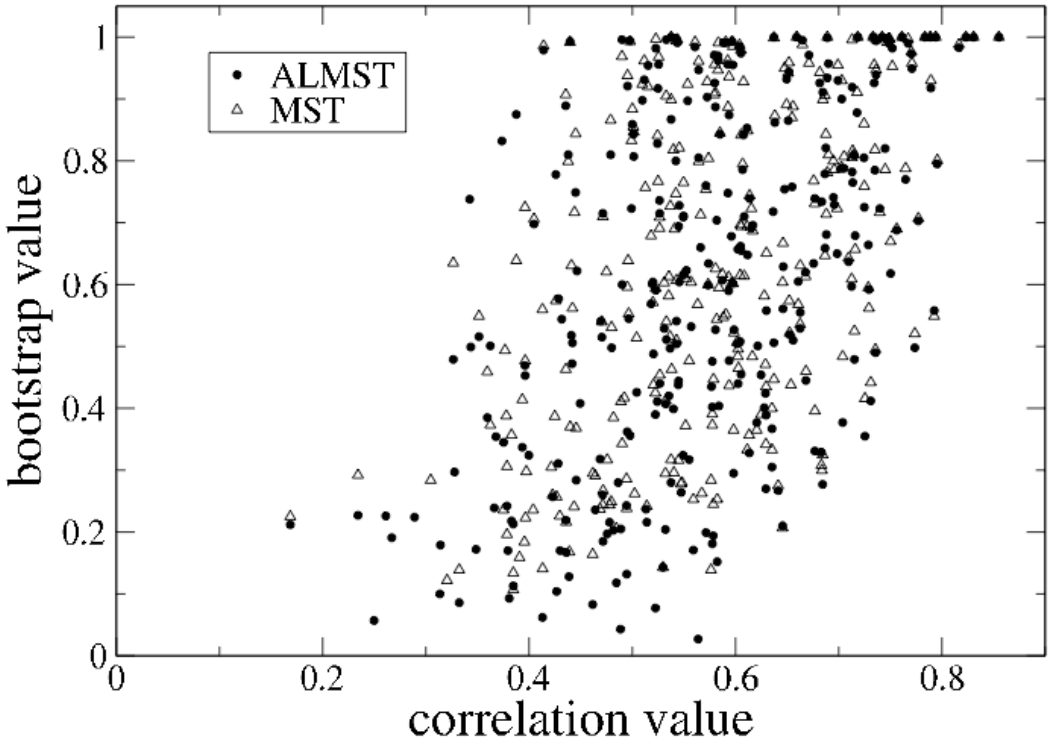
Which are the determinants of the bootstrap value of a link?

- The value of the correlation between the two elements;

- The topology of the correlation based network.

OCS

Bootstrap vs correlation



$N = 300$ (NYSE)

daily returns

2001 - 2003

$T = 748$

For Gaussian series: $\sigma_\rho = \frac{1-\rho^2}{\sqrt{T-3}}$

OCS

Information and stability of a filtering procedure. How to quantify it?

The Kullback-Leibler divergence

OCS

Kullback-Leibler distance

We propose to use the Kullback-Leibler divergence to quantify the performance of different filtering procedures of the correlation matrix

$$K(p,q) = E_p\left[\log\left(\frac{p}{q}\right)\right] \quad , \qquad \text{where } p \text{ and } q \text{ are pdf's.}$$

For **multivariate normally distributed** random variables we obtain:

$$K(P(\Sigma_1, X), P(\Sigma_2, X)) = \frac{1}{2}\left[\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + \mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right) - n\right] = K(\Sigma_1, \Sigma_2)$$

Minimizing the Kullback-Leibler distance is equivalent to maximizing the likelihood in the maximum likelihood factor analysis.

M. Tumminello, F. Lillo and R.N.M., Phys. Rev. E **76**, 031123 (2007)

OCS

Expectation values

$$E\left[K(\Sigma, S_1)\right] = \frac{1}{2}\left\{ n \log\left(\frac{2}{T}\right) + \sum_{p=T-n+1}^{T}\left[\frac{\Gamma^I(p/2)}{\Gamma(p/2)}\right] + \frac{n(n+1)}{T-n-1}\right\}$$

$$E\left[K(S_1, \Sigma)\right] = \frac{1}{2}\left\{ n \log\left(\frac{T}{2}\right) - \sum_{p=T-n+1}^{T}\left[\frac{\Gamma^I(p/2)}{\Gamma(p/2)}\right]\right\}$$

$$E\left[K(S_1, S_2)\right] = \frac{1}{2}\frac{n(n+1)}{T-n-1}$$

where Σ is the model correlation matrix of the system while S_1 and S_2 are two sample correlation matrices obtained from two independent realizations each of length T .

The three expectation values are independent from Σ, i.e they do not depend from the underlying model !

OCS

Kullback vs Frobenius

- The expectation values of Frobenius distance are model dependent, e.g. for a system of $n=2$ Gaussian random variables with correlation coefficient ϱ it is

$$\mathrm{E}\left[F(\Sigma,\mathbf{S})\right] = \mathrm{E}\left[\sqrt{\mathrm{tr}\left[(\Sigma-\mathbf{S})(\Sigma-\mathbf{S})^{\mathrm{T}}\right]}\right] = \frac{2}{\sqrt{\pi T}}\left(1-\rho^2\right)$$

where Σ is the model correlation matrix of the system while $\mathbf{S}$ is a sample correlation matrix obtained from a realization of length T.

OCS

Kullback-Leibler divergence for non Gaussian random variables

The Kullback-Leibler divergence can also be analytically calculated random variables following a **multivariate Student's t-distribution**:

$$P_{\Sigma}(\mathbf{x},\mu) = \frac{\Gamma\left(\frac{n+\mu}{2}\right)}{\Gamma\left(\frac{\mu}{2}\right)\sqrt{(\mu\,\pi)^n\,|\Sigma|}}\,\frac{1}{\left[1+\frac{1}{\mu}\,\tilde{\mathbf{x}}\,\Sigma^{-1}\mathbf{x}\right]^{\frac{n+\mu}{2}}}$$

If $\frac{\mu}{n} << 1$ then :

$$K(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2) = \frac{1}{2}\left[\log\left(\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|}\right) + n\,\log\left(\frac{\mathrm{tr}\left(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\right)}{n}\right)\right]$$

G. Biroli, J.-P. Bouchaud, M. Potters, Acta Phys. Pol. B 38, 4009 (2007)

OCS

Gaussian vs Student's t-distribution

$$K_G(\Sigma_1,\Sigma_2) = \frac{1}{2}\left[\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + \mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right) - n\right]$$

$$K_S(\Sigma_1,\Sigma_2) = \frac{1}{2}\left\{\log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + n\log\left[\frac{\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right)}{n}\right]\right\}$$

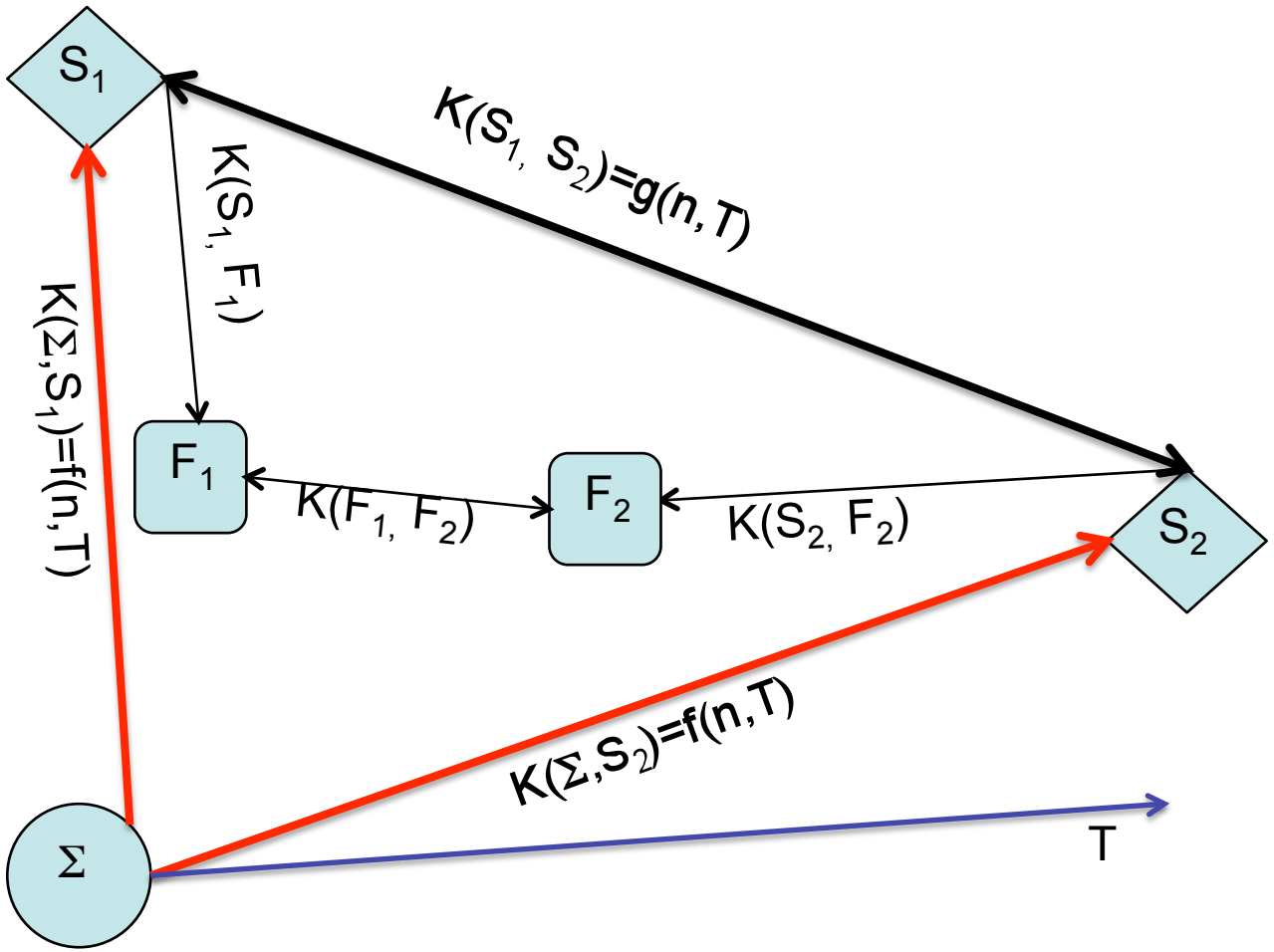
$$\text{If } \Sigma_1 \cong \Sigma_2 \Rightarrow\ K_G(\Sigma_1,\Sigma_2) \cong K_S(\Sigma_1,\Sigma_2)$$

OCS

The Kullback-Leibler divergence in the analysis of models and empirical data

OCS

Comparison of filtering procedures



S_1 and **S_2** are sample correlation matrices estimated from independent realizations or bootstrap-replicas of the system.

F_1 and **F_2** are matrices filtered from S_1 and S_2 respectively.

Σ is the true correlation matrix of the system.

OCS

Accuracy of filtering procedures (Models)

TABLE I. Average value of the Kullback-Leibler distance between the correlation matrix of the model and the correlation matrix filtered from the sample. For each case, average and standard deviations are obtained from 1000 realizations or bootstrap replicas of the system ($n=100$, $T=748$).

Models	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{\mathrm{ALCA}})\rangle$	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{\mathrm{SLCA}})\rangle$	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{B})\rangle$	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{S})\rangle$
Block diagonal	0.15±0.01	0.57±0.04	0.84±0.03	1.50±0.05
HNFM	0.22±0.02	0.33±0.05	1.99±0.07	2.15±0.08
Block diagonal (n.d.)	3.56±0.02	4.36±0.07	3.74±0.06	4.34±0.09
HNFM (n.d.)	3.38±0.02	3.85±0.08	4.54±0.08	5.0±0.1

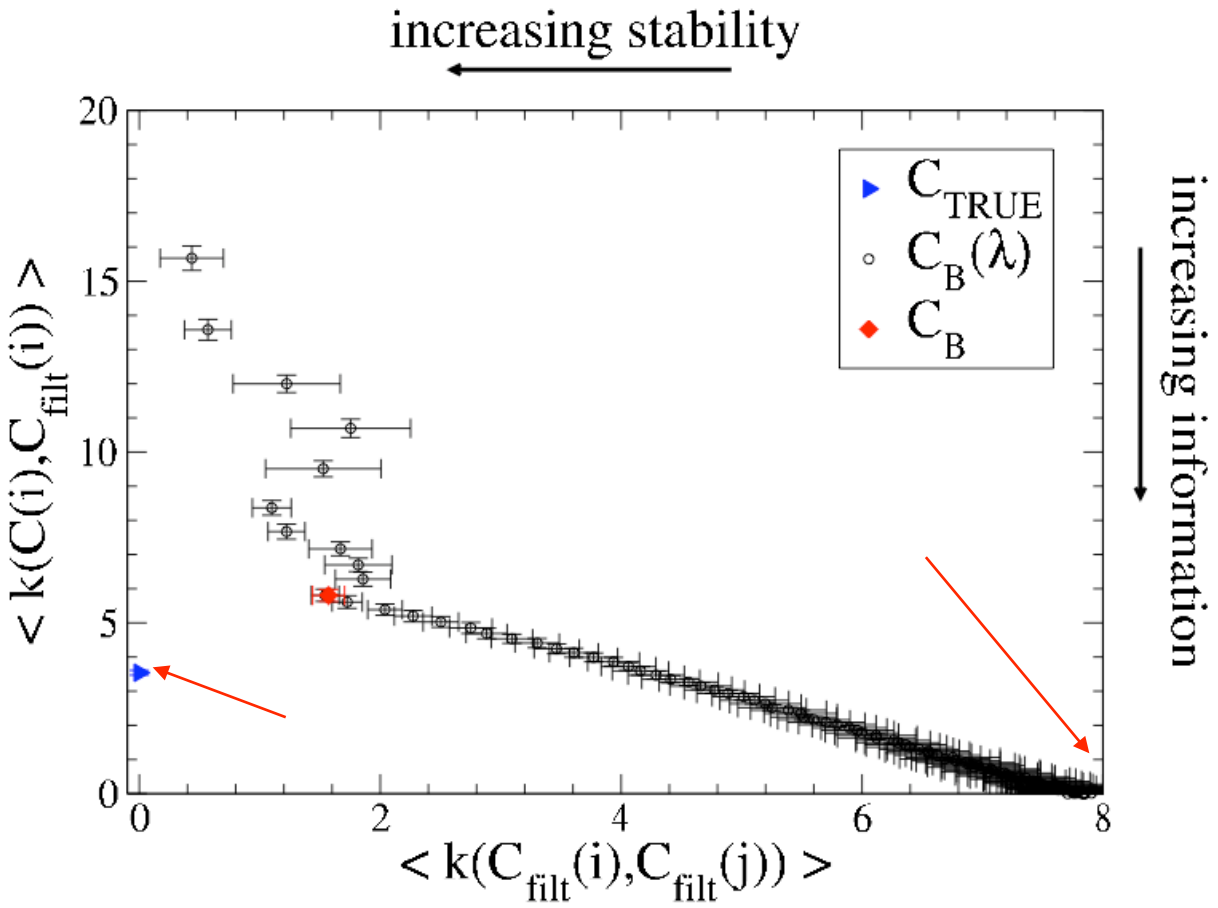
TABLE II. The same as in Table I but with $T=7480$.

Models	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{\mathrm{ALCA}})\rangle$	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{\mathrm{SLCA}})\rangle$	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{B})\rangle$	$\langle K(\mathbf{\Sigma},\mathbf{C}_i^{S})\rangle$
Block diagonal	0.015±0.001	0.105±0.006	0.162±0.006	0.70±0.01
HNFM	0.023±0.002	0.032±0.005	0.986±0.007	1.44±0.07
Block diagonal (n.d.)	3.418±0.004	3.94±0.02	2.95±0.02	3.41±0.02
HNFM (n.d.)	3.174±0.008	3.52±0.02	2.54±0.04	4.66±0.09

M. Tumminello, F. Lillo and R.N.M., Phys. Rev. E **76**, 031123 (2007)

OCS

Comparison of filtered correlation matrices



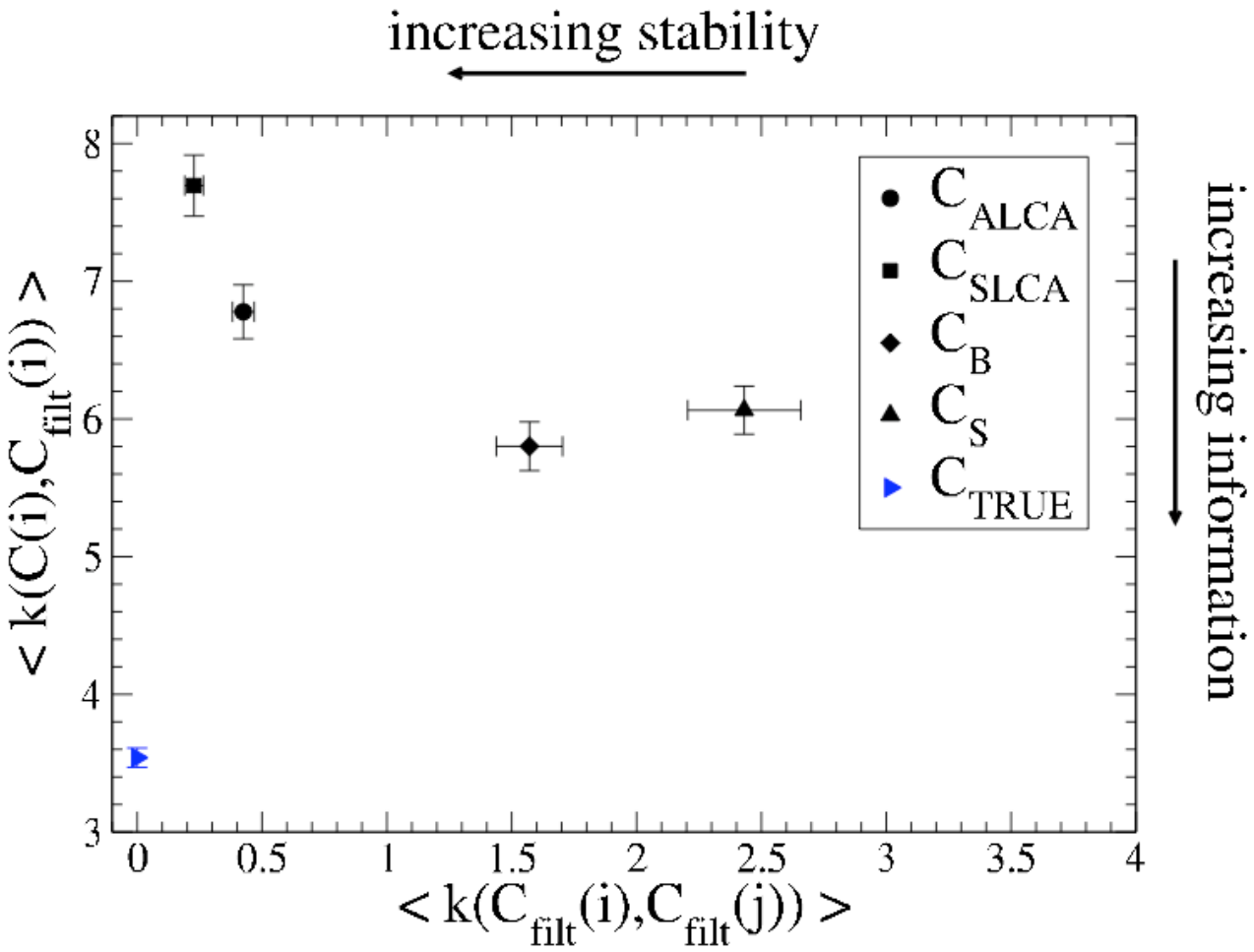
Block diagonal model with 12 factors.

N=100,
T=748.

Gaussian random Variables.

OCS

Comparison of filtered correlation matrices



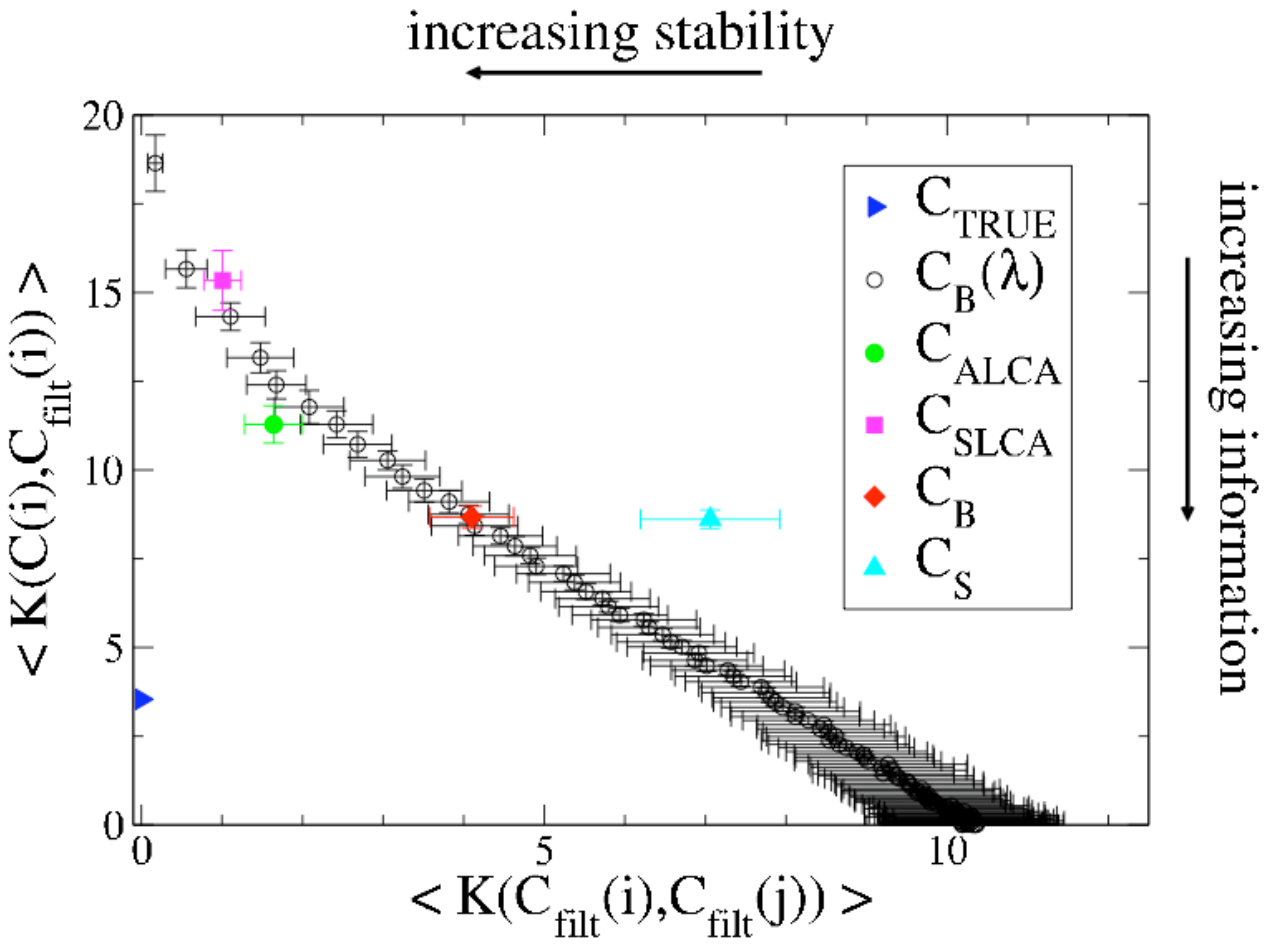
Block diagonal model with 12 factors.

N=100,
T=748.

Gaussian random Variables.

OCS

Comparison of filtered correlation matrices (real data)



$N = 100$ (NYSE)

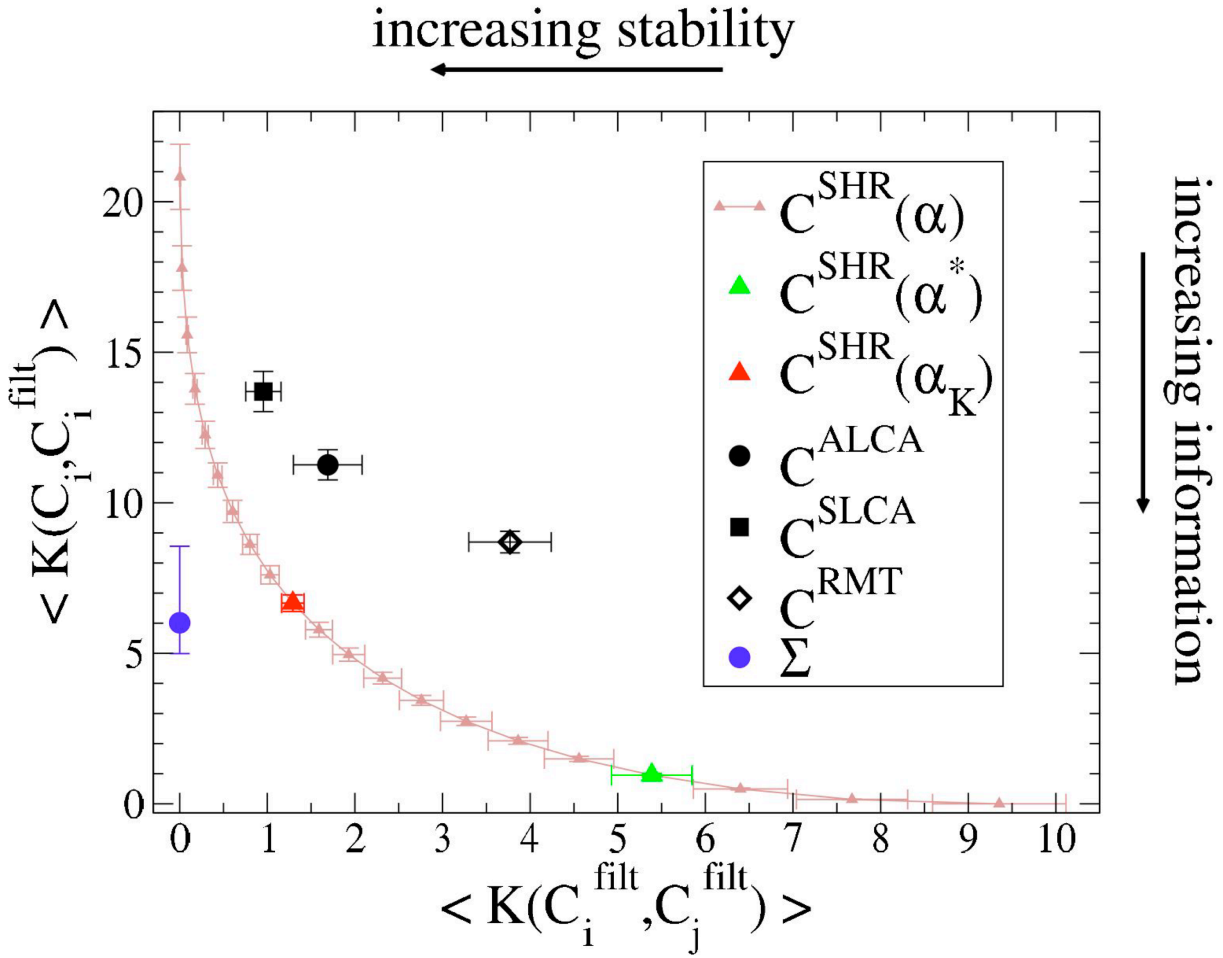
daily returns

2001 - 2003

$T = 748$

OCS

Comparison of filtered correlation matrices (real data, Student's t assumption)



$N = 100$ (NYSE)

daily returns

2001 - 2003

$T = 748$

M. Tumminello, F. Lillo, R.N. Mantegna, ACTA PHYSICA POLONICA B 38, 4079-4088 (2007)

OCS

The Hierarchically Nested Factor Model

A factor is associated to each node

$$x_i(t) = \sum_{\alpha_h \in G(i)} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - \sum_{\alpha_h \in G(i)} \gamma_{\alpha_h}^2}\, \varepsilon_i(t)$$

α_h-th factor

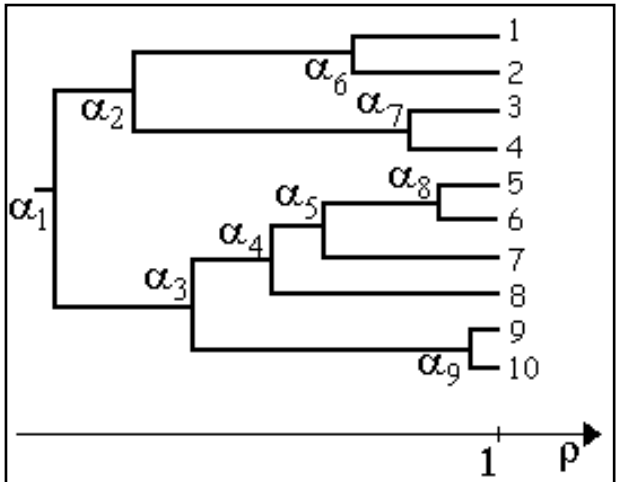
Idiosyncratic term

$$\gamma_{\alpha_h} = \sqrt{\rho_{\alpha_h} - \rho_{g(\alpha_h)}}; \gamma_{\alpha_1} = \sqrt{\rho_{\alpha_1}}$$

The model explains $C^{<} = (\rho_{ij}^{<})$

$$\langle x_i \cdot x_j \rangle = \sum_{\alpha_h \in G(i) \cap G(j)} \gamma_{\alpha_h}^2 = \rho_{\alpha_k} = \rho_{ij}^{<}$$

e.g. $\langle x_1 \cdot x_4 \rangle = \gamma_{\alpha_2}^2 + \gamma_{\alpha_1}^2 = \rho_{\alpha_2} - \rho_{\alpha_1} + \rho_{\alpha_1} = \rho_{\alpha_2}$



$G(i)$ = Pedigree of element i,

e.g. $G(9) = \{\alpha_1, \alpha_3, \alpha_9\}$

$g(\alpha_h)$ = Parent of node α_h,

e.g. $g(\alpha_7) = \alpha_2$

M. Tumminello, F. Lillo, R.N. Mantegna, Hierarchically nested factor model from multivariate data, EPL 78 (3), Art. No. 30006 (2007).

A simple hierarchically nested factor model

$$x_i(t) = \gamma_0 f^{(0)}(t) + \gamma_1 f^{(1)}(t) + \sqrt{1 - \gamma_0^2 - \gamma_1^2}\, \varepsilon_i(t) \quad \text{for } i \le n_1$$

$$x_i(t) = \gamma_0 f^{(0)}(t) + \gamma_2 f^{(2)}(t) + \sqrt{1 - \gamma_0^2 - \gamma_2^2}\, \varepsilon_i(t) \quad \text{for } n_1 < i \le N$$

$$\mathbf{C} = \begin{bmatrix} \rho_1 & & \\ & & \\ & \rho_2 & \\ \rho_M & & \end{bmatrix} \begin{matrix} \} n_1 \\ \\ \} n_2 = N - n_1 \end{matrix}$$

$$\rho_1 = \gamma_0^2 + \gamma_1^2$$

$$\rho_2 = \gamma_0^2 + \gamma_2^2$$

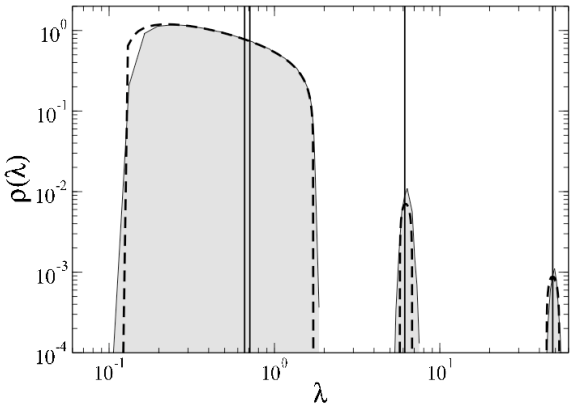
$$\rho_M = \gamma_0^2$$

OCS

Spectral Analysis

2 large eigenvalues

$$\lambda^{\pm} = \frac{1}{2}\left[2 + q_{+} \pm (q_{-}^{\,2} + 4\, n_1\, n_2\, \rho_M^{\,2})^{1/2}\right]$$



2 corresponding eigenvectors

$$\vec{z^{\pm}} = (u^{\pm} \ldots . u^{\pm}, v^{\pm}, \ldots v^{\pm})$$

where $u^{\pm} = 1/\sqrt{2\, n_1 \left[1 + y^2 \mp y\sqrt{1 + y^2}\right]}$, $v^{\pm} = \pm 1/\sqrt{2\, n_2 \left[1 + y^2 \pm y\sqrt{1 + y^2}\right]}$

$$q_{\pm} = (n_1 - 1)\rho_1 \pm (n_2 - 1)\rho_2 \quad \text{and} \quad y = q_{-}/(4\, n_1\, n_2\, \rho_M^{\,2})^{1/2}$$

PCA is not able to reconstruct the true model and/or to give insights about its hierarchical features

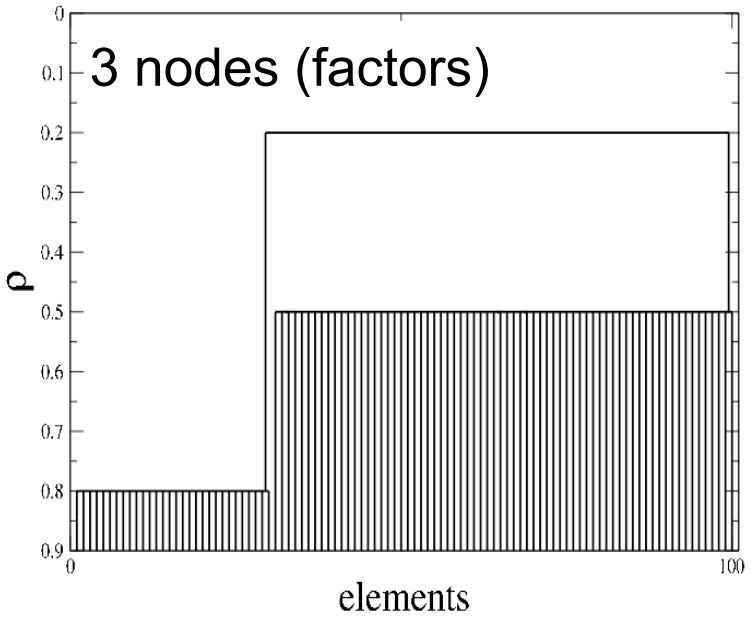
OCS

The hierarchically nested factor model allows to simulate the system. We use hierarchical clustering to investigate the simulations so that we can estimate the ability of hierarchical clustering to detect a hierarchically nested system.

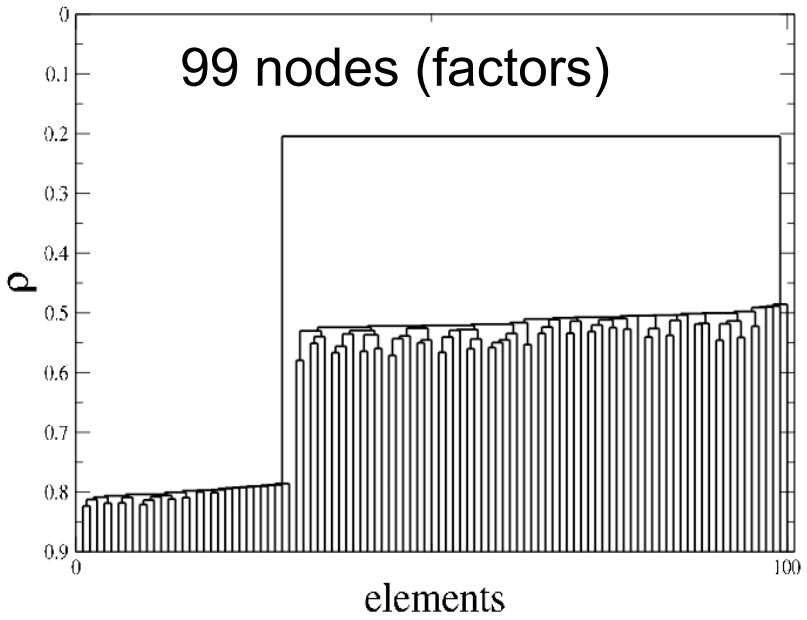
A problem of the hierarchical clustering method: by hierarchical clustering we always detects a n-1 factor hierarchically nested factor model

A solution: Evaluation of node statistical uncertainty and node reduction

Hierarchical tree of the model



Hierarchical tree reconstruction



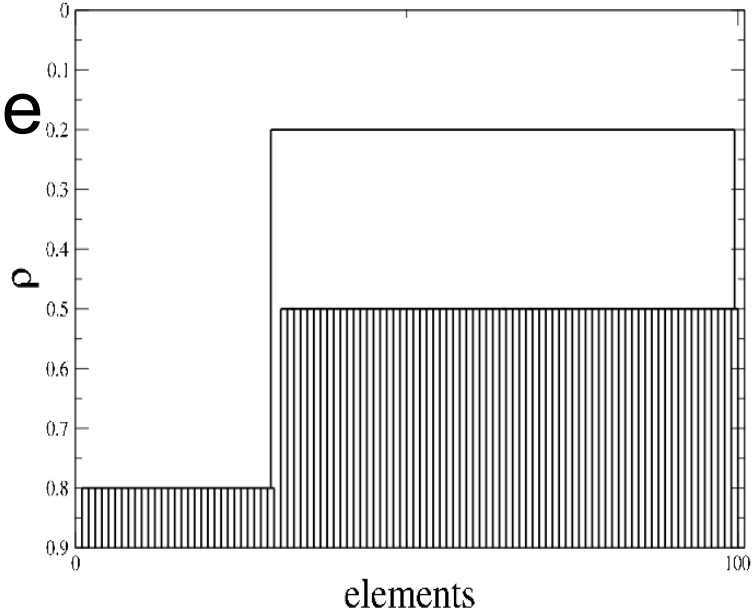
OCS

Self-consistent node-factor reduction

- Select a bootstrap value threshold b_t .

- For each node α_k :

  If $b(\alpha_k) < b_t$ then merge the node α_k with his first ancestor α_q (in the path to the root) such that $b(\alpha_q) \geq b_t$.



- How to chose b_t ?

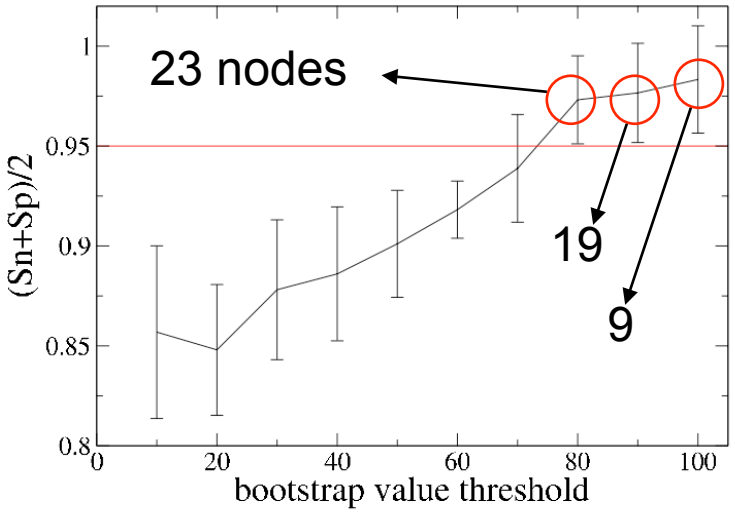
In a self-consistent way!

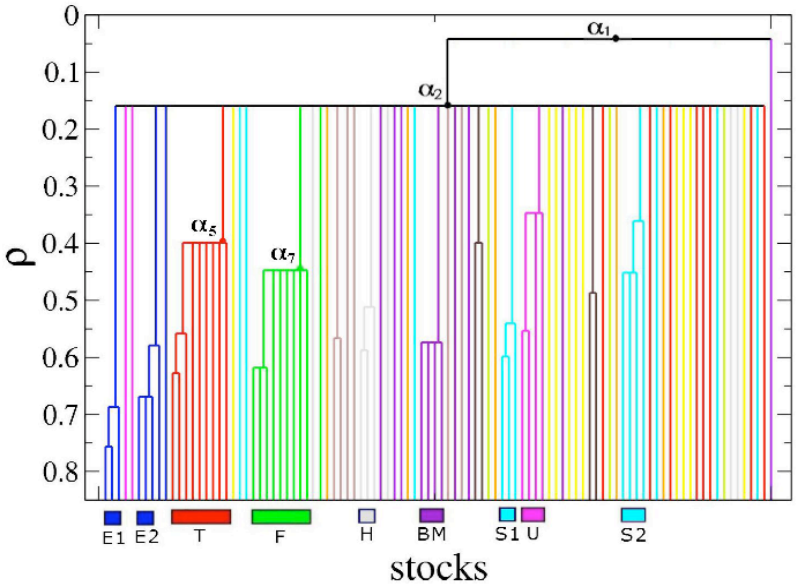
Hierarchically nested factor model correctly detects the model when b_t>0.70

OCS

Node reduction for an empirical system

Daily return of 100 stocks traded at NYSE in the time period 1/1995-12/1998 (T=1011)





Sn = sensitivity; Sp = specificity

$$S_n = \frac{TP}{TP + FN} \qquad S_p = \frac{TN}{TN + FP}$$

OCS

Interpretation of factors

Hierarchically nested factor model associated with the reduced dendrogram of 23 nodes. Here we made explicit the equations valid for stocks belonging to the Technology and Financial sectors.

Financial Factor

$$x_i^F(t) = \gamma_{\alpha_7} f^{(\alpha_7)}(t) + \sum_{h=1}^{2} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - (\gamma_{\alpha_7}^2 + \sum_{h=1}^{2} \gamma_{\alpha_h}^2)}\ \epsilon_i(t),$$

$$x_j^T(t) = \gamma_{\alpha_5} f^{(\alpha_5)}(t) + \sum_{h=1}^{2} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - (\gamma_{\alpha_5}^2 + \sum_{h=1}^{2} \gamma_{\alpha_h}^2)}\ \epsilon_j(t).$$

Technology Factor

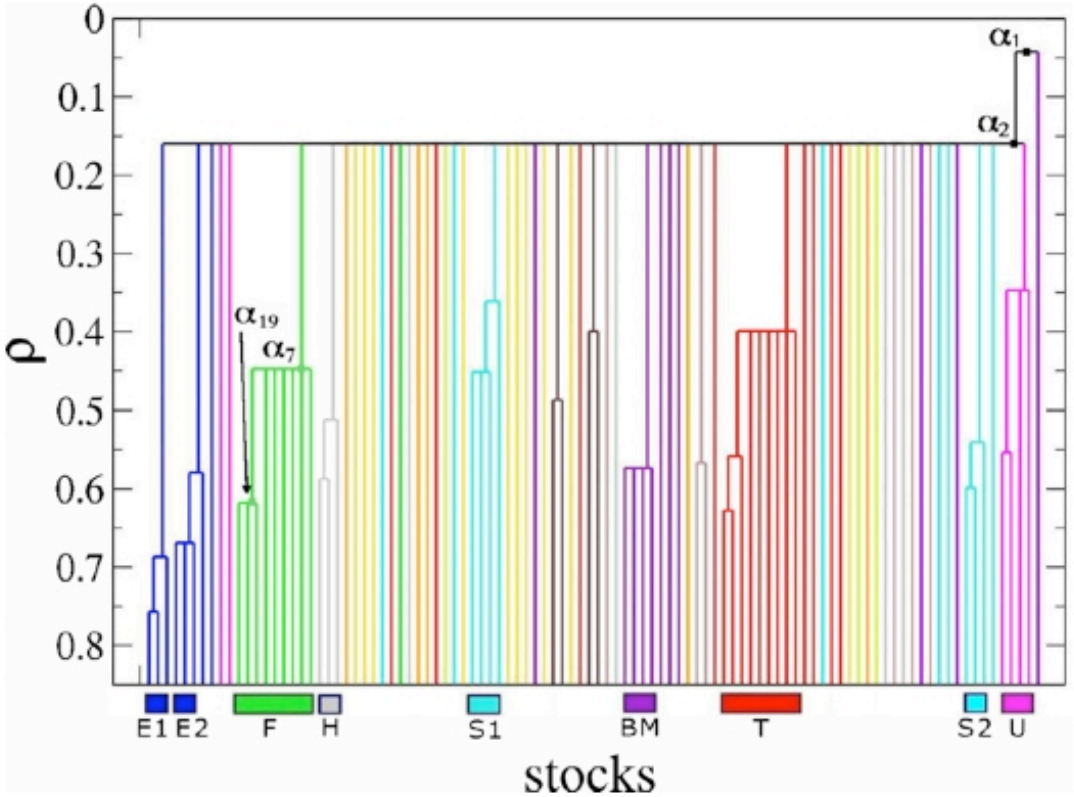
OCS

Financial Factor

$$x_i^F(t) = \gamma_{\alpha_7} f^{(\alpha_7)}(t) + \sum_{h=1}^{2} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - (\gamma_{\alpha_7}^2 + \sum_{h=1}^{2} \gamma_{\alpha_h}^2)}\ \epsilon_i(t),$$

$$x_j^T(t) = \gamma_{\alpha_5} f^{(\alpha_5)}(t) + \sum_{h=1}^{2} \gamma_{\alpha_h} f^{(\alpha_h)}(t) + \sqrt{1 - (\gamma_{\alpha_5}^2 + \sum_{h=1}^{2} \gamma_{\alpha_h}^2)}\ \epsilon_j(t).$$

Technology Factor



Financial companies:
BAC, JPM, MER, AGC, AOG, AXP, ONE, WFC, USB

OCS

- Correlation based networks are informative about complex systems;

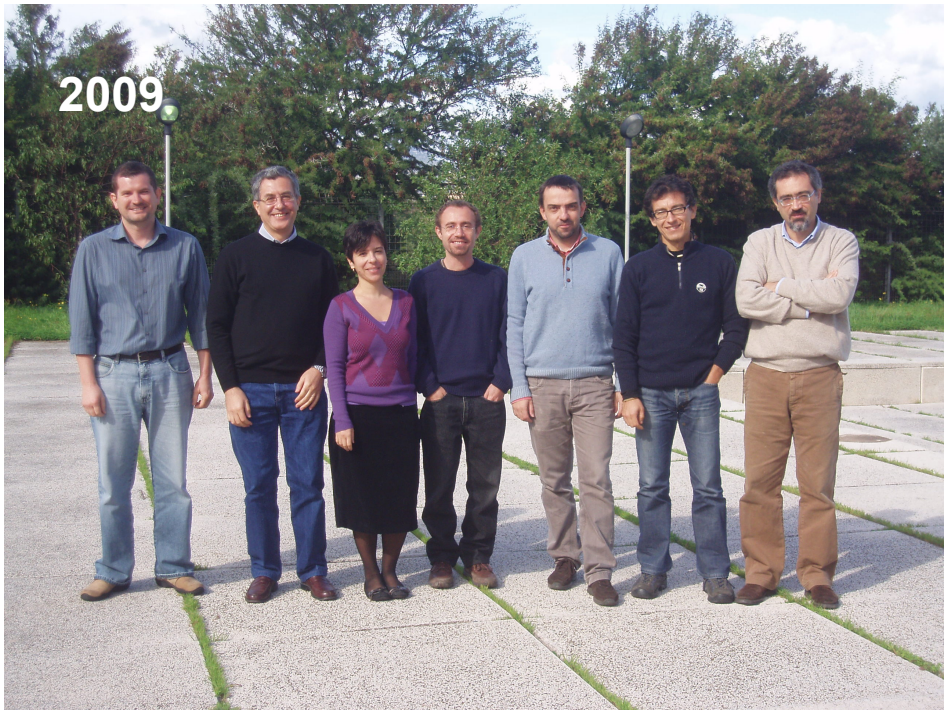
- Bootstrap can be used to statistically validate hierarchical trees and correlation based networks;

- The stability and amount of information of a filtering procedure of a correlation matrix can be quantified by using the Kullback-Leibler distance.

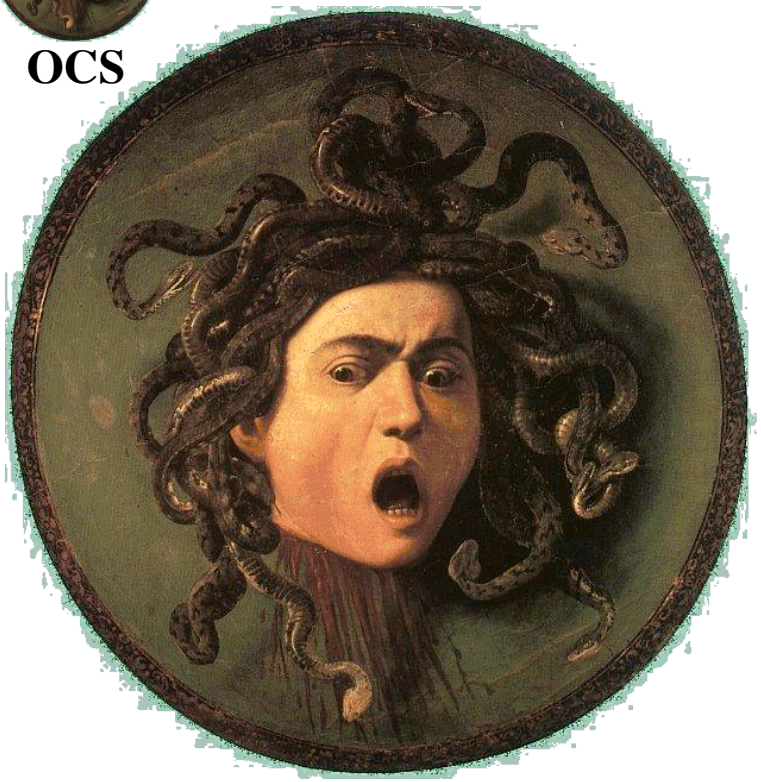
OCS

***O**bservatory of **C**omplex **S**ystems*

http://ocs.unipa.it



OCS



Michelangelo Merisi
da Caravaggio
Head of Medusa (1598)
Uffizi gallery

"The only hero able to cut off Medusa's head is Perseus, who flies with winged sandals. ..... . To cut off Medusa's head without being turned to stone, Perseus supports himself on the very lightest of things, the winds and the clouds, and fixes his gaze upon what can be revealed only by indirect vision, an image caught in a mirror. I am immediately tempted to see this myth as an allegory on the poet's relationship to the world, a lesson in the method to follow when writing."

Italo Calvino, Six Memos for the Next MillenniumVintage Books,
Random House, New York 1988

OCS website: http://ocs.unipa.it