A Random Matrix Theory based analysis of stocks of markets from different countries

Ricardo Coelho

School of Physics
Trinity College Dublin

Palermo - Presentation, 22^{nd} September 2007

Outline

Outline

Introduction

Definitions

- Logarithmic Return
- Correlations
- Random Matrix Theory
- Distances

Outline

Introduction

Definitions

- Logarithmic Return
- Correlations
- Random Matrix Theory
- Distances

Results

- Markets are clustered geographically
- For some markets stocks are clustered in sectors
- Stocks from different markets cluster in Industrial/Geographical location?

Outline

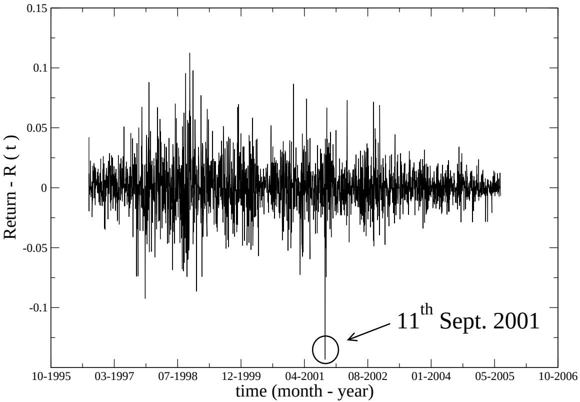
Introduction

We studied different sets of data.

- Different World Indices
  - Equity markets of 53 different countries
  - Wednesday closing price (weekly returns)
  - From 8^{th} January 1997 until 1^{st} February 2006
- More than 6000 stocks from markets around the world
  - Daily closing price
  - From 30^{th} December 1994 until 1^{st} January 2007

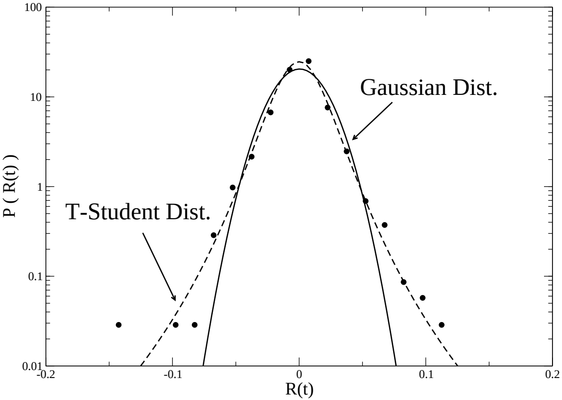
Log-return of stocks shows extreme events (e.g. HSBC)

$$R_i(t) = \ln P_i(t) - \ln P_i(t-1) \tag{1}$$



Distribution of the Log-return is non-Gaussian

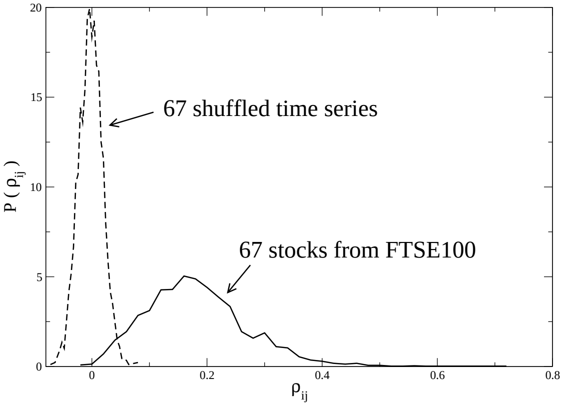
$k \sim 2.90$ and $q \sim 1.34$



Correlations between stocks i and j

$$\rho_{ij} = \frac{\langle \mathbf{R}_i \mathbf{R}_j \rangle - \langle \mathbf{R}_i \rangle \langle \mathbf{R}_j \rangle}{\sqrt{\left(\langle \mathbf{R}_i^2 \rangle - \langle \mathbf{R}_i \rangle^2\right) \times \left(\langle \mathbf{R}_j^2 \rangle - \langle \mathbf{R}_j \rangle^2\right)}} \quad (2)$$

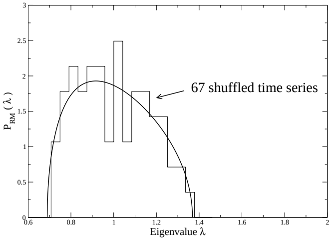
- ρ_{ij} form a symmetric $N \times N$ matrix; $-1 \leq \rho_{ij} \leq 1$; $\rho_{ii} = 1$.

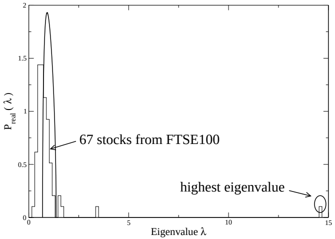


Matrix of correlations shows non-randomness

Random Matrix Theory for $N \to \infty$ and $T \to \infty$, where $Q = T/N$ is fixed and bigger than 1.

$$P_{RM}(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} \qquad (3)$$

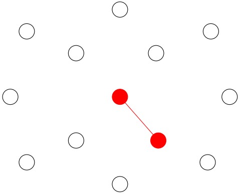




Distance between stocks i and j

$$d_{ij} = \sqrt{2(1-\rho_{ij})} \tag{4}$$

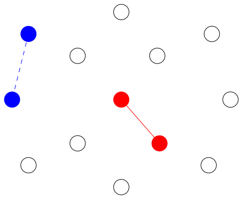
- $0 \le d_{ij} \le 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Choose the minimum distance between 2 stocks - one link between these 2.



Distance between stocks i and j

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{4}$$

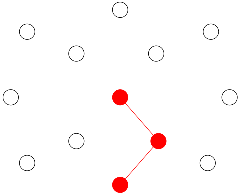
- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Choose the minimum distance between 2 stocks - one link between these 2.
- Find the next minimum distance. If the 2 stocks are not linked or both linked, choose the next minimum distance. If just 1 of the stocks has already links, link them.



Distance between stocks i and j

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{4}$$

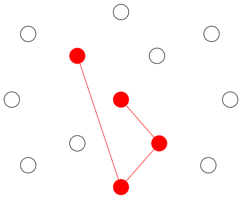
- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Choose the minimum distance between 2 stocks - one link between these 2.
- Find the next minimum distance. If the 2 stocks are not linked or both linked, choose the next minimum distance. If just 1 of the stocks has already links, link them.



Distance between stocks i and j

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{4}$$

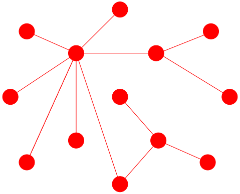
- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Choose the minimum distance between 2 stocks - one link between these 2.
- Find the next minimum distance. If the 2 stocks are not linked or both linked, choose the next minimum distance. If just 1 of the stocks has already links, link them.



Distance between stocks i and j

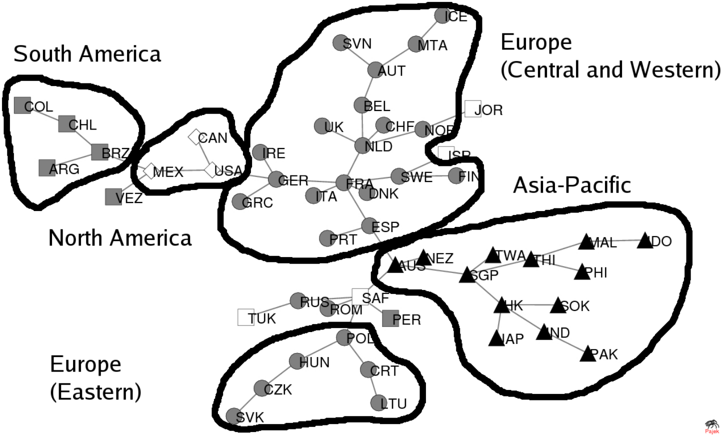
$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{4}$$

- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Choose the minimum distance between 2 stocks - one link between these 2.
- Find the next minimum distance. If the 2 stocks are not linked or both linked, choose the next minimum distance. If just 1 of the stocks has already links, link them.



Markets are organised by geographical location

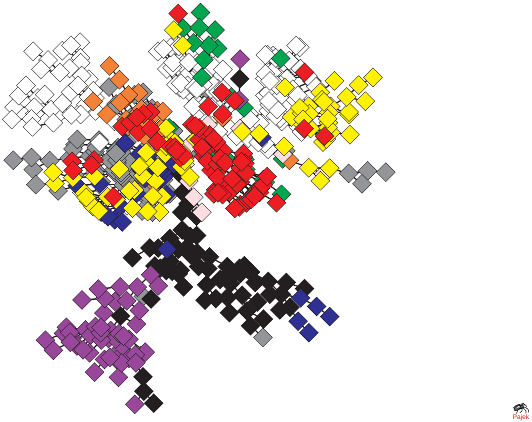
Full time series, $T = 475$ weeks, for the 53 World Indices.



Coelho et al., Physica A 376 (2007) 455-466

Stocks cluster in industrial sectors (NYSE stocks)

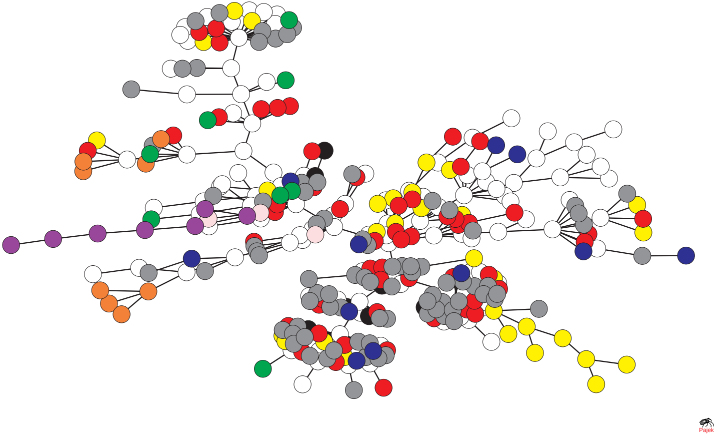
Full time series, $T = 3127$ days, for 617 NYSE stocks.
Oil & Gas - black / Basic Materials - blue / Industrials - gray / Consumer Goods - yellow / Health Care - green
Consumer Services - red / Telecommunications - pink / Utilities - purple / Financials - white / Technology - orange



Clustering depends on the portfolio (LON stocks)

Full time series, $T = 3127$ days, for 322 LON stocks.

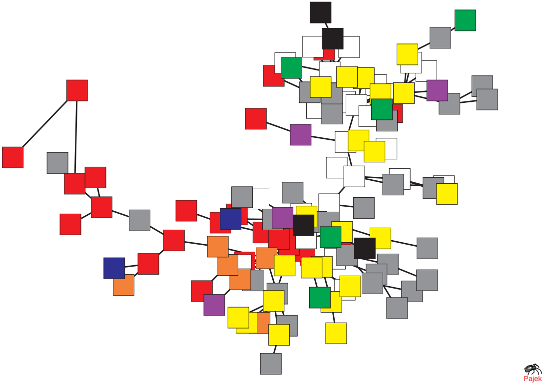
Oil & Gas - black / Basic Materials - blue / Industrials - gray / Consumer Goods - yellow / Health Care - green
Consumer Services - red / Telecommunications - pink / Utilities - purple / Financials - white / Technology - orange



Clustering depends on the portfolio (PAR stocks)

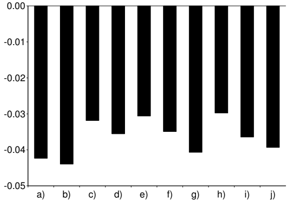
Full time series, $T = 3127$ days, for 125 PAR stocks.

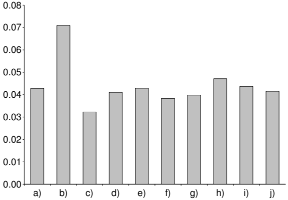
Oil & Gas - black / Basic Materials - blue / Industrials - gray / Consumer Goods - yellow / Health Care - green
Consumer Services - red / Utilities - purple / Financials - white / Technology - orange



Eigenvector elements of the highest eigenvalue

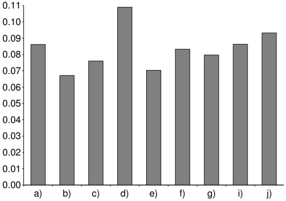
Industrials - a / Financials - b / Health Care - c / Technology - d / Oil & Gas - e / Utilities - f
Basic Materials - g / Telecommunications - h / Consumer Goods - i / Consumer Services - j





NYSE

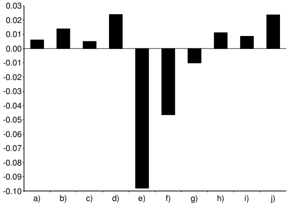
LON



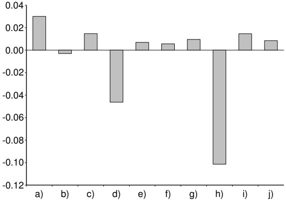
PAR

Eigenvector elements of the 2^{nd} highest eigenvalue

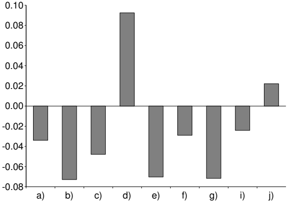
Industrials - a / Financials - b / Health Care - c / Technology - d / Oil & Gas - e / Utilities - f
Basic Materials - g / Telecommunications - h / Consumer Goods - i / Consumer Services - j



NYSE



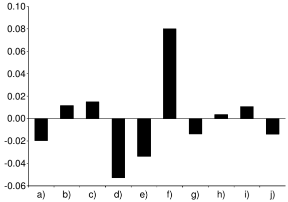
LON



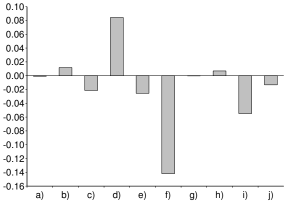
PAR

Eigenvector elements of the 3^{rd} highest eigenvalue

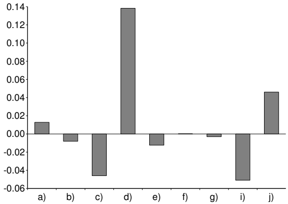
Industrials - a / Financials - b / Health Care - c / Technology - d / Oil & Gas - e / Utilities - f
Basic Materials - g / Telecommunications - h / Consumer Goods - i / Consumer Services - j



NYSE



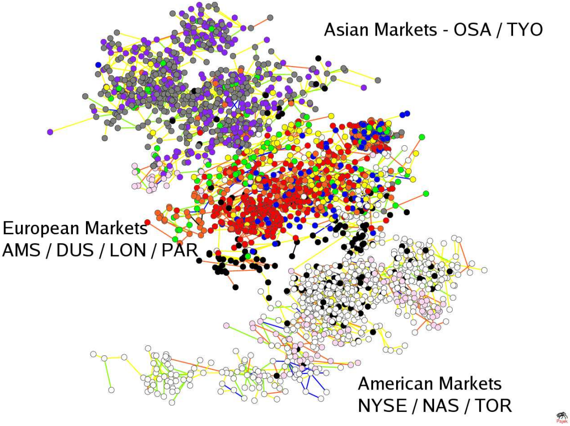
LON



PAR

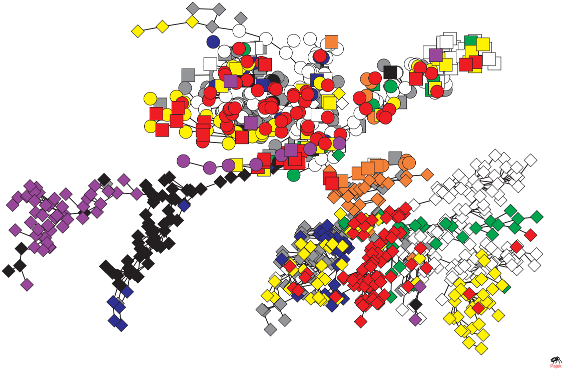
What about stocks from different countries???

Full time series, $T = 3127$ days, for 2500 stocks from different markets.



Stocks from European and American markets segregate

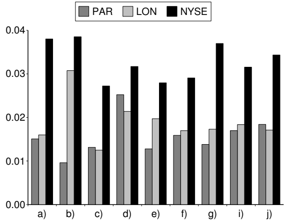
1052 stocks from NYSE (◊), LON (○) and PAR (□). Oil & Gas - black / Basic Materials - blue / Industrials - gray / Consumer Goods - yellow / Health Care - green Consumer Services - red / Utilities - purple / Financials - white / Technology - orange



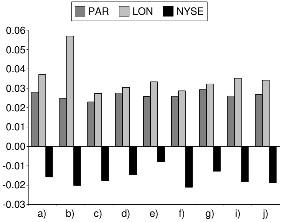
back

Stocks from European and American markets segregate

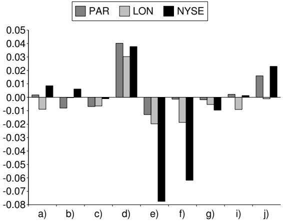
Industrials - a / Financials - b / Health Care - c / Technology - d / Oil & Gas - e / Utilities - f
Basic Materials - g / Consumer Goods - i / Consumer Services - j



Highest



2^{nd} highest



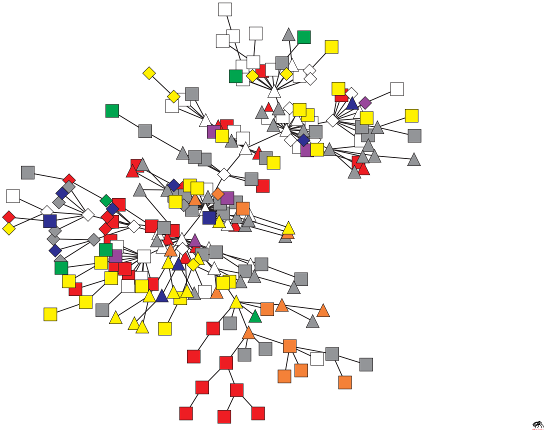
3^{rd} highest

back

Stocks from European markets mixed together

235 stocks from PAR (□), AMS (△) and BRU (◊).

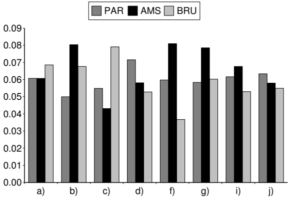
Basic Materials - blue / Industrials - gray / Consumer Goods - yellow / Health Care - green
Consumer Services - red / Utilities - purple / Financials - white / Technology - orange



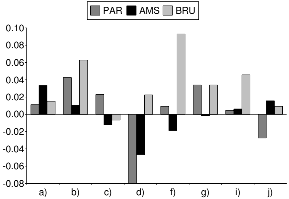
back

Stocks from European markets mixed together

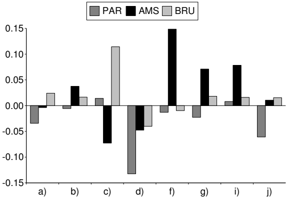
Industrials - a / Financials - b / Health Care - c / Technology - d / Utilities - f / Basic Materials - g
Consumer Goods - i / Consumer Services - j



Highest



2^{nd} highest



3^{rd} highest

back

What about simulations of three diferent markets?

Using information from the real time series of returns, construct random time series for all stocks:

$$r_i(t) = \alpha_i + \beta_i R^{M_j}(t) + \epsilon_i(t) \tag{5}$$

where

$$R^{M_j}(t) = \sum_{i \in M_j} R_i(t) \tag{6}$$

The correlations between market returns is given by:

PAR		AMS		BRU			
Real	Rand	Real	Rand	Real	Rand		
1.0	0.97	0.83		0.77		Real	PAR
	1.0		0.78		0.70	Rand	
		1.0	0.96	0.77		Real	AMS
			1.0		0.69	Rand	
				1.0	0.94	Real	BRU
					1.0	Rand	

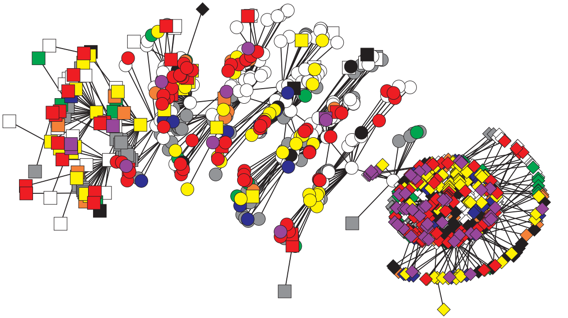
NYSE is weakly correlated with PAR and LON

And for the other three markets the correlations are given by:

PAR		LON		NYSE			
Real	Rand	Real	Rand	Real	Rand		
1.0	0.97	0.68		0.28		Real	PAR
	1.0		0.66		0.27	Rand	
		1.0	0.99	0.31		Real	LON
			1.0		0.30	Rand	
				1.0	0.99	Real	NYSE
					1.0	Rand	

Simulation shows segregation in terms of market

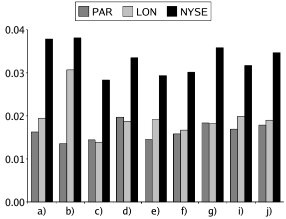
1052 stocks from NYSE (◊), LON (○) and PAR (□). Oil & Gas - black / Basic Materials - blue / Industrials - gray / Consumer Goods - yellow / Health Care - green Consumer Services - red / Utilities - purple / Financials - white / Technology - orange



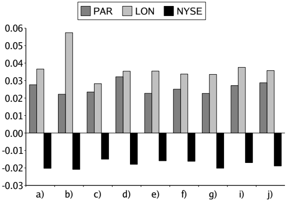
▸ To Real MST

Simulation shows segregation in terms of market

Industrials - a / Financials - b / Health Care - c / Technology - d / Oil & Gas - e / Utilities - f
Basic Materials - g / Consumer Goods - i / Consumer Services - j



Highest



2^{nd} highest

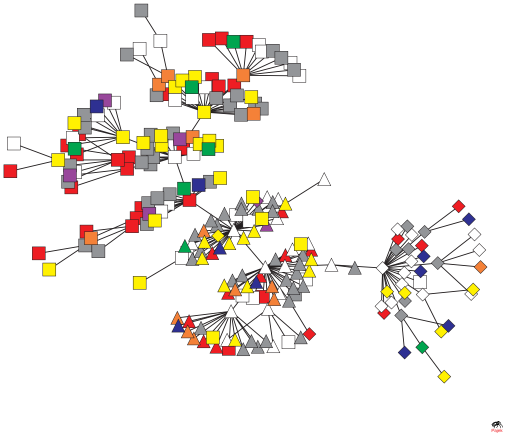
3^{rd} highest

To Real Eigenvectors

Simulation shows similar behaviour for any random market

235 stocks from PAR (□), AMS (△) and BRU (◊).

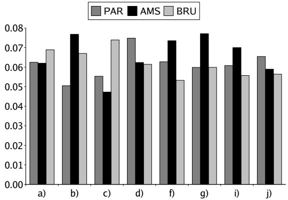
Basic Materials - blue / Industrials - gray / Consumer Goods - yellow / Health Care - green
Consumer Services - red / Utilities - purple / Financials - white / Technology - orange



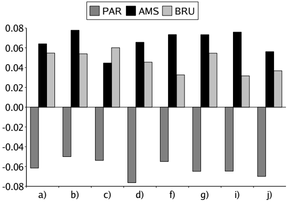
To Real MST

Simulation shows similar behaviour for any random market

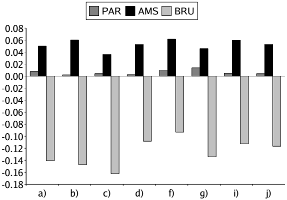
Industrials - a / Financials - b / Health Care - c / Technology - d / Utilities - f / Basic Materials - g
Consumer Goods - i / Consumer Services - j



Highest



2^{nd} highest

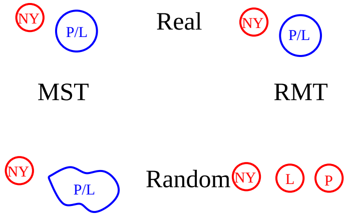


3^{rd} highest

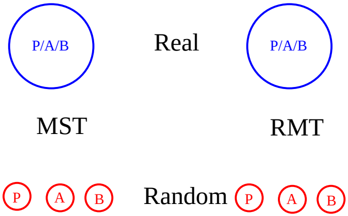
To Real Eigenvectors

Conclusions

For NYSE / LON / PAR:



For PAR / AMS / BRU:



Conclusions

- Two different methodologies of analysing large amount of data not always give same results
- MST show different clusters (Industrial sectors / Geography)
- The choice of the stocks/market can influence the clustering in sectors
- The stocks cluster in market before clustering in sectors
- The correlation between indices influence this clustering
- Less correlation between different countries shows better segregation between stocks of each country
- Random time series from different sources (markets) show the same results
- Random time series show segregation between different markets

Econophysics group of the Trinity College Dublin:

- Prof. Peter Richmond
- Dr. Stefan Hutzler
- Ricardo Coelho
- Joseph Barry

Collaborations:

- Prof. Brian Lucey (Business School of Trinity College Dublin)
- Prof. Claire Gilmore (McGowan School of Business of King's College, Pennsylvania)

Part of this work was published in Physica A:

`Coelho et al., Physica A 373 (2007) 615-626`

`Coelho et al., Physica A 376 (2007) 455-466`

Acknowledgments

- SFI - Science Foundation Ireland
- FCT - Fundação para a Ciência e a Tecnologia