Minimal Spanning Trees (MST) analysis of random time series from different distributions

Ricardo Coelho

School of Physics
Trinity College Dublin

COST P10, Vienna, September 2006

Outline

Outline

1. Introduction
2. Definitions
   - Price
   - Logarithmic Return
   - Correlations
   - Distances
3. Minimal Spanning Trees
   - Real MST
   - Random MST
   - Correlations
   - Distances
   - Degree Distribution
4. Summary

Outline

Outline

Introduction

Econophysics group of the Trinity College Dublin:

- Prof. Peter Richmond
- Prof. Stefan Hutzler
- Dr. Przemyslaw Repetowicz
- Shane Huston

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 can be found in the Physics arXiv:

physics/0601189

physics/0607022

We studied two different sets of data.

- A portfolio from the London Stock Exchange FTSE100 index:
  - 67 stocks
  - Daily closing price
  - From 2^{nd} August 1996 until 27^{th} June 2005
- Different World Indices
  - 53 countries' equity markets
  - Wednesday closing price (weekly returns)
  - From 8^{th} January 1997 until 1^{st} February 2006

We studied two different sets of data.

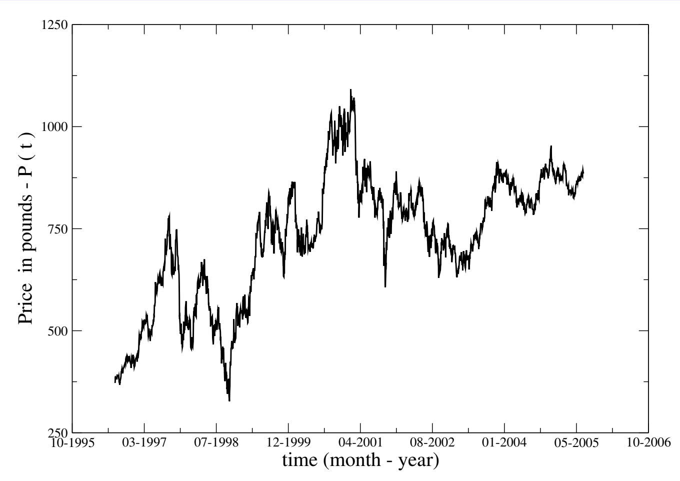
- A portfolio from the London Stock Exchange FTSE100 index:
  - 67 stocks
  - Daily closing price
  - From 2^{nd} August 1996 until 27^{th} June 2005
- Different World Indices
  - 53 countries' equity markets
  - Wednesday closing price (weekly returns)
  - From 8^{th} January 1997 until 1^{st} February 2006

We studied two different sets of data.

- A portfolio from the London Stock Exchange FTSE100 index:
  - 67 stocks
  - Daily closing price
  - From 2nd August 1996 until 27th June 2005
- Different World Indices
  - 53 countries' equity markets
  - Wednesday closing price (weekly returns)
  - From 8th January 1997 until 1st February 2006

Outline

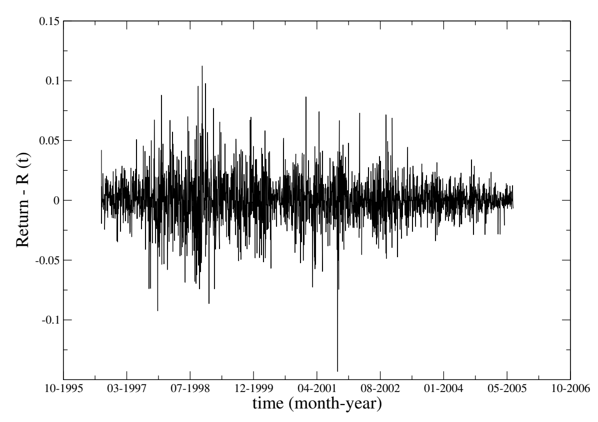
Price of a stock from the FTSE100 (HSBC)



Outline

Log-return of a stock from the FTSE100 (HSBC)

$$R_i(t) = \ln P_i(t) - \ln P_i(t-1) \quad (1)$$



Outline

Correlation between stocks i and j

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

- $\langle \cdots \rangle$ is an average over time $\frac{1}{T}\sum_{t'=t}^{t+T-1} \cdots$, t is the first day, T is the length of our time series.
- $-1 \leq \rho_{ij} \leq 1$.
- ρ_{ij} form a symmetric $N \times N$ matrix. $\rho_{ii} = 1$.

Correlation between stocks *i* and *j*

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

- $\langle \cdots \rangle$ is an average over time $\frac{1}{T}\sum_{t'=t}^{t+T-1} \cdots$, t is the first day, T is the length of our time series.
- $-1 \leq \rho_{ij} \leq 1$.
- ρ_{ij} form a symmetric $N \times N$ matrix. $\rho_{ii} = 1$.

Correlation between stocks i and j

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

- $\langle \cdots \rangle$ is an average over time $\frac{1}{T}\sum_{t'=t}^{t+T-1} \cdots$, t is the first day, T is the length of our time series.
- $-1 \leq \rho_{ij} \leq 1$.
- ρ_{ij} form a symmetric $N \times N$ matrix. $\rho_{ii} = 1$.

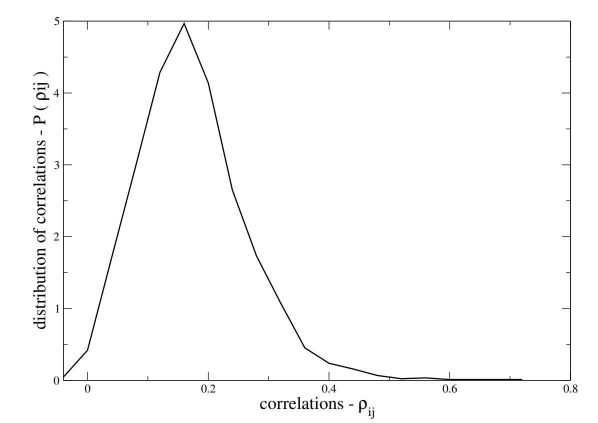
Correlation between stocks i and j

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

- $\langle \cdots \rangle$ is an average over time $\frac{1}{T}\sum_{t'=t}^{t+T-1} \cdots$, t is the first day, T is the length of our time series.
- $-1 \leq \rho_{ij} \leq 1$.
- ρ_{ij} form a symmetric $N \times N$ matrix. $\rho_{ii} = 1$.

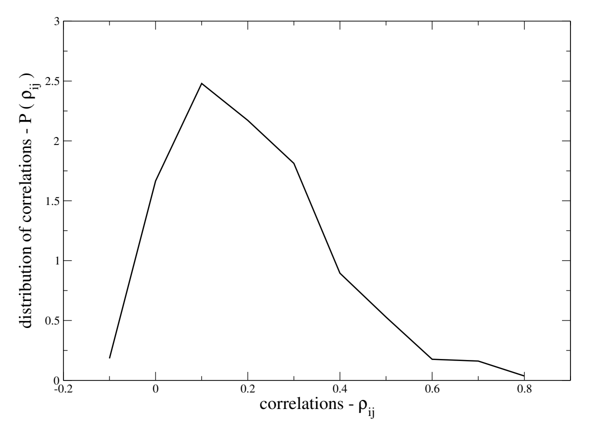
Distribution of the correlations

Full time series, $T = 2322$ days, for the FTSE100 stocks.



Distribution of the correlations

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



Moments of the correlations

- Mean

$$\overline{\rho} = \frac{2}{N(N-1)} \sum_{i<j} \rho_{ij} \tag{3}$$

- Variance

$$\lambda_2 = \frac{2}{N(N-1)} \sum_{i<j} (\rho_{ij} - \overline{\rho})^2 \tag{4}$$

- Skewness

$$\lambda_3 = \frac{2}{N(N-1)\lambda_2^{3/2}} \sum_{i<j} (\rho_{ij} - \overline{\rho})^3 \tag{5}$$

- Kurtosis

$$\lambda_4 = \frac{2}{N(N-1)\lambda_2^{2}} \sum_{i<j} (\rho_{ij} - \overline{\rho})^4 \tag{6}$$

Moments of the correlations

- Mean

$$\overline{\rho} = \frac{2}{N(N-1)} \sum_{i<j} \rho_{ij} \tag{3}$$

- Variance

$$\lambda_2 = \frac{2}{N(N-1)} \sum_{i<j} (\rho_{ij} - \overline{\rho})^2 \tag{4}$$

- Skewness

$$\lambda_3 = \frac{2}{N(N-1)\lambda_2^{3/2}} \sum_{i<j} (\rho_{ij} - \overline{\rho})^3 \tag{5}$$

- Kurtosis

$$\lambda_4 = \frac{2}{N(N-1)\lambda_2^{2}} \sum_{i<j} (\rho_{ij} - \overline{\rho})^4 \tag{6}$$

Moments of the correlations

- Mean

$$\overline{\rho} = \frac{2}{N(N-1)} \sum_{i<j} \rho_{ij} \tag{3}$$

- Variance

$$\lambda_2 = \frac{2}{N(N-1)} \sum_{i<j} (\rho_{ij} - \overline{\rho})^2 \tag{4}$$

- Skewness

$$\lambda_3 = \frac{2}{N(N-1)\lambda_2^{3/2}} \sum_{i<j} (\rho_{ij} - \overline{\rho})^3 \tag{5}$$

- Kurtosis

$$\lambda_4 = \frac{2}{N(N-1)\lambda_2^{2}} \sum_{i<j} (\rho_{ij} - \overline{\rho})^4 \tag{6}$$

Moments of the correlations

- Mean

$$\bar{\rho} = \frac{2}{N(N-1)} \sum_{i<j} \rho_{ij} \tag{3}$$

- Variance

$$\lambda_2 = \frac{2}{N(N-1)} \sum_{i<j} (\rho_{ij} - \bar{\rho})^2 \tag{4}$$

- Skewness

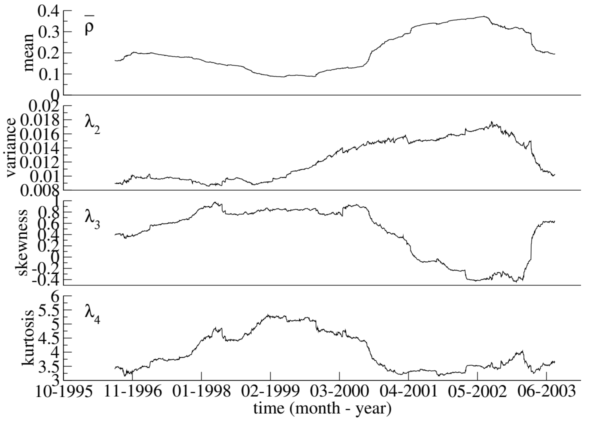
$$\lambda_3 = \frac{2}{N(N-1)\lambda_2^{3/2}} \sum_{i<j} (\rho_{ij} - \bar{\rho})^3 \tag{5}$$

- Kurtosis

$$\lambda_4 = \frac{2}{N(N-1)\lambda_2^{2}} \sum_{i<j} (\rho_{ij} - \bar{\rho})^4 \tag{6}$$

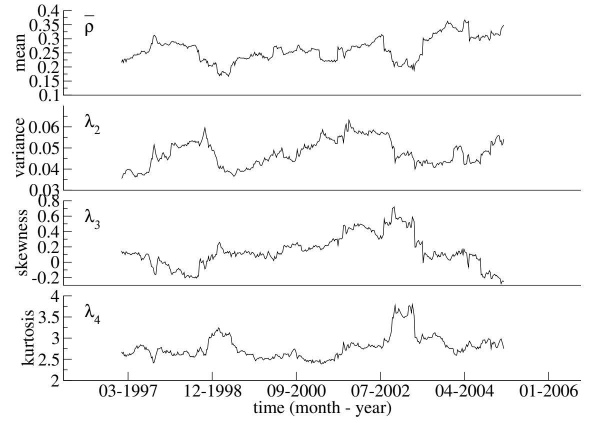
Correlations and anti-correlations between moments

FTSE100, $T = 500$ days, $\delta T = 1$ day.



Correlations and anti-correlations between moments

World Indices, $T = 52$ weeks, $\delta T = 1$ week.



Outline

Distance between stocks i and j

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{7}$$

- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Analyse the properties of these networks.

Distance between stocks *i* and *j*

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{7}$$

- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Analyse the properties of these networks.

Distance between stocks i and j

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{7}$$

- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Analyse the properties of these networks.

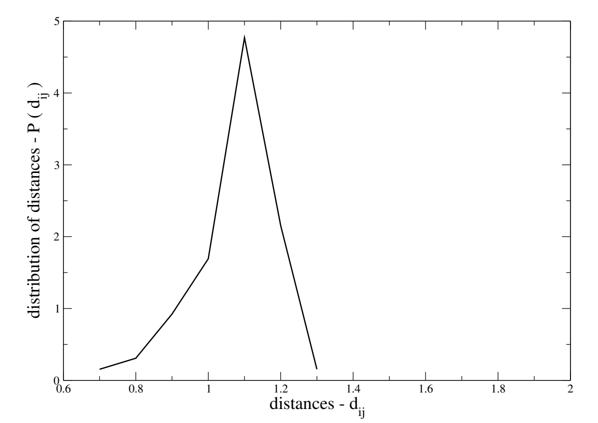
Distance between stocks i and j

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \tag{7}$$

- $0 \leq d_{ij} \leq 2$, small values imply strong correlations.
- From distances we construct the Minimal Spanning Tree (MST).
- Analyse the properties of these networks.

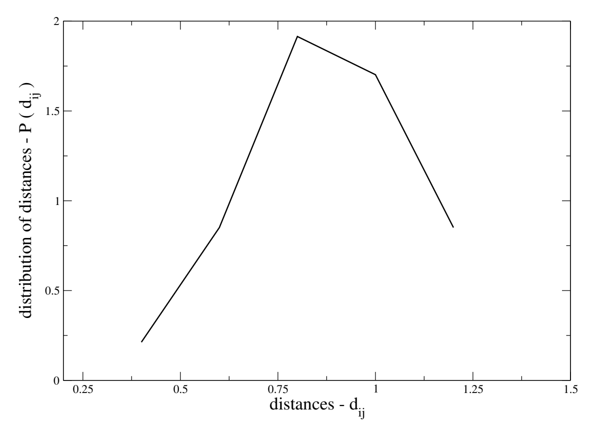
Distribution of the distances

Full time series, $T = 2322$ days, for the FTSE100 stocks.



Distribution of the distances

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



Moments of the distances

- Mean (Normalized tree length)

$$L = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} d_{ij} \tag{8}$$

- Variance

$$\nu_2 = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^2 \tag{9}$$

- Skewness

$$\nu_3 = \frac{1}{(N-1)\nu_2^{3/2}} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^3 \tag{10}$$

- Kurtosis

$$\nu_4 = \frac{1}{(N-1)\nu_2^2} \sum (d_{ij} - L)^4 \tag{11}$$

Moments of the distances

- Mean (Normalized tree length)

$$L = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} d_{ij} \quad (8)$$

- Variance

$$\nu_2 = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^2 \quad (9)$$

- Skewness

$$\nu_3 = \frac{1}{(N-1)\nu_2^{3/2}} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^3 \quad (10)$$

- Kurtosis

$$\nu_4 = \frac{1}{(N-1)\nu_2^{2}} \sum (d_{ij} - L)^4 \quad (11)$$

Moments of the distances

- Mean (Normalized tree length)

$$L = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} d_{ij} \tag{8}$$

- Variance

$$\nu_2 = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^2 \tag{9}$$

- Skewness

$$\nu_3 = \frac{1}{(N-1)\nu_2^{3/2}} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^3 \tag{10}$$

- Kurtosis

$$\nu_4 = \frac{1}{(N-1)\nu_2^{2}} \sum (d_{ij} - L)^4 \tag{11}$$

Moments of the distances

- Mean (Normalized tree length)

$$L = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} d_{ij} \tag{8}$$

- Variance

$$\nu_2 = \frac{1}{N-1} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^2 \tag{9}$$

- Skewness

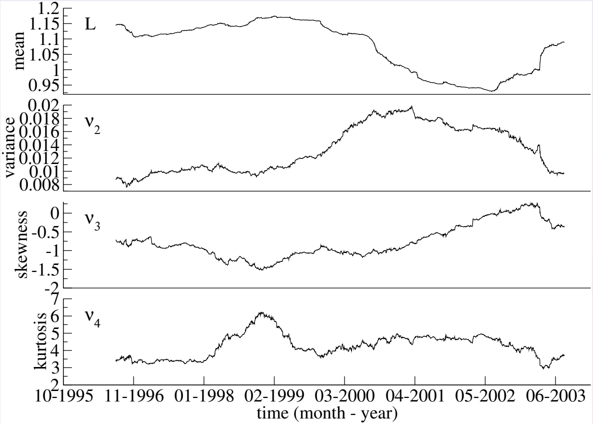
$$\nu_3 = \frac{1}{(N-1)\nu_2^{3/2}} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^3 \tag{10}$$

- Kurtosis

$$\nu_4 = \frac{1}{(N-1)\nu_2^2} \sum_{d_{ij} \in \Theta} (d_{ij} - L)^4 \tag{11}$$

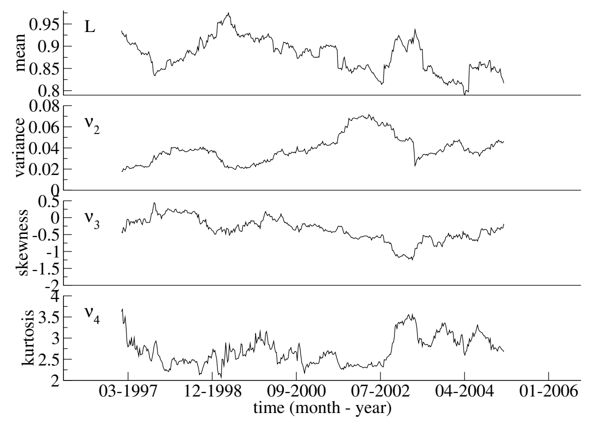
Correlations and anti-correlations between moments

FTSE100, $T = 500$ days, $\delta T = 1$ day.



Correlations and anti-correlations between moments

World Indices, $T = 52$ weeks, $\delta T = 1$ week.

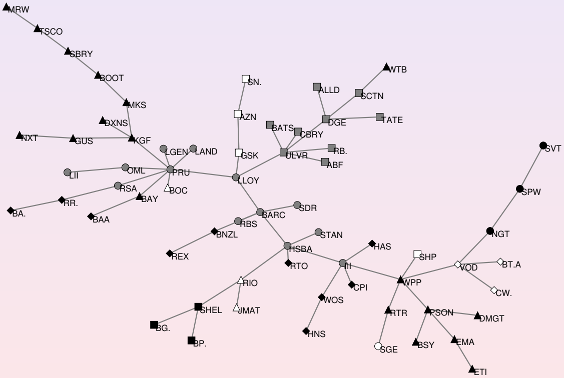


Outline

1. Introduction
2. Definitions
   - Price
   - Logarithmic Return
   - Correlations
   - Distances
3. Minimal Spanning Trees
   - Real MST
   - Random MST
   - Correlations
   - Distances
   - Degree Distribution
4. Summary

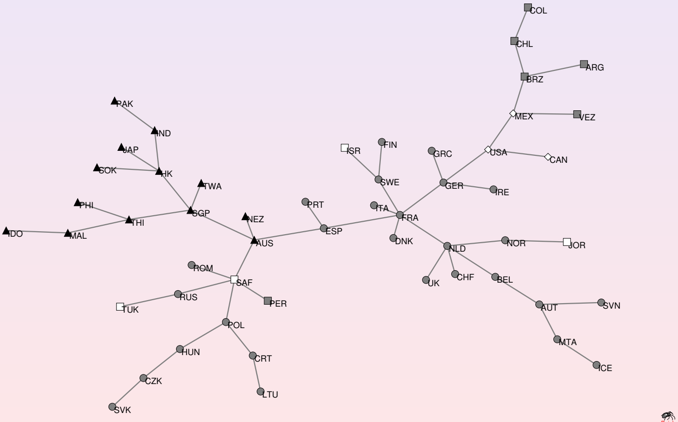
Stocks cluster in industrial sectors (ICB classification)

Full time series, $T = 2322$ days, for the FTSE100 stocks.



Markets organized by geographical location

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



Outline

Can we mimic real MST from random data?

- Returns as random variables from a Gaussian distribution

$$r_i(t) = \epsilon_i(t) \tag{12}$$

- Returns as random variables from a T-student distribution

$$r_i(t) = \gamma_i(t) \tag{13}$$

- Returns from a *Market Model* with Gaussian noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \epsilon_i(t) \tag{14}$$

- Returns from a *Market Model* with T-student noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \gamma_i(t) \tag{15}$$

Can we mimic real MST from random data?

- Returns as random variables from a Gaussian distribution

$$r_i(t) = \epsilon_i(t) \tag{12}$$

- Returns as random variables from a T-student distribution

$$r_i(t) = \gamma_i(t) \tag{13}$$

- Returns from a *Market Model* with Gaussian noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \epsilon_i(t) \tag{14}$$

- Returns from a *Market Model* with T-student noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \gamma_i(t) \tag{15}$$

Can we mimic real MST from random data?

- Returns as random variables from a Gaussian distribution

$$r_i(t) = \epsilon_i(t) \tag{12}$$

- Returns as random variables from a T-student distribution

$$r_i(t) = \gamma_i(t) \tag{13}$$

- Returns from a *Market Model* with Gaussian noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \epsilon_i(t) \tag{14}$$

- Returns from a *Market Model* with T-student noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \gamma_i(t) \tag{15}$$

Can we mimic real MST from random data?

- Returns as random variables from a Gaussian distribution

$$r_i(t) = \epsilon_i(t) \tag{12}$$

- Returns as random variables from a T-student distribution

$$r_i(t) = \gamma_i(t) \tag{13}$$

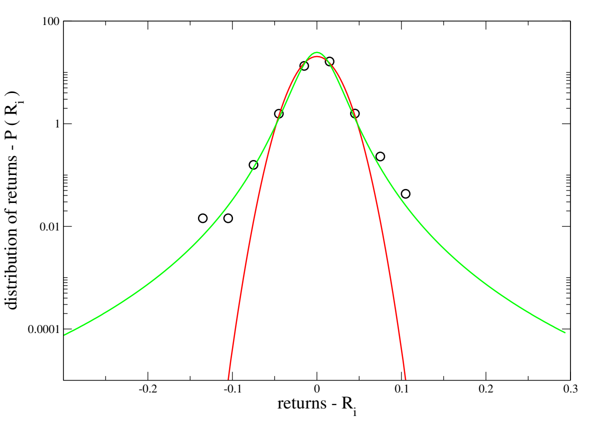
- Returns from a *Market Model* with Gaussian noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \epsilon_i(t) \tag{14}$$

- Returns from a *Market Model* with T-student noise

$$r_i(t) = \alpha_i + \beta_i R_m(t) + \gamma_i(t) \tag{15}$$

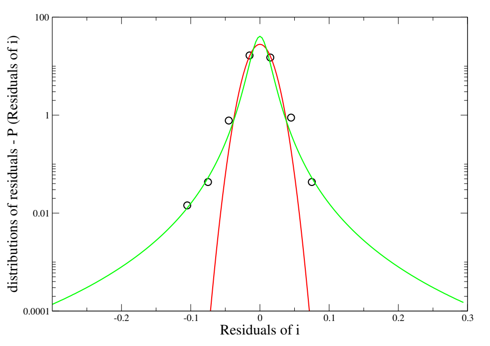
Distribution of the Log-return (HSBC)



Distribution of the Residuals (HSBC)

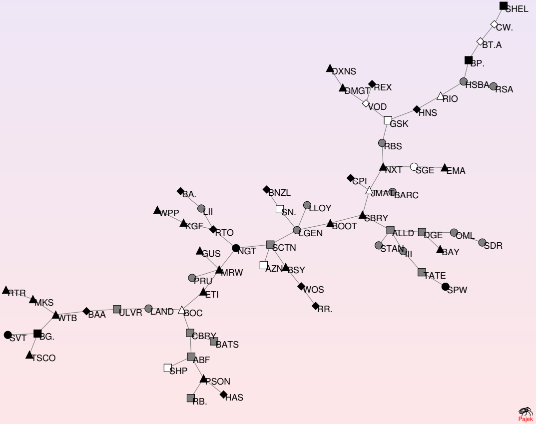
Residuals of a stock is the difference between the returns and the values estimated using the *Market Model*:

$$Res_i(t) = R_i(t) - \alpha_i - \beta_i R_M(t) \quad (16)$$



Random Model - Gaussian

Full time series, $T = 2322$ days (pseudo-FTSE100 stocks).

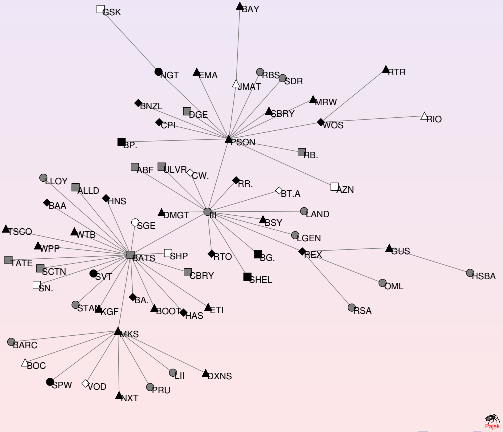


Random Model - T-student

Full time series, $T = 2322$ days (pseudo-FTSE100 stocks).

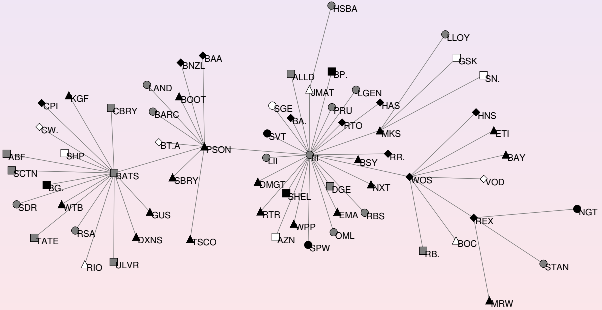
Market Model - Gaussian

Full time series, $T = 2322$ days (pseudo-FTSE100 stocks).



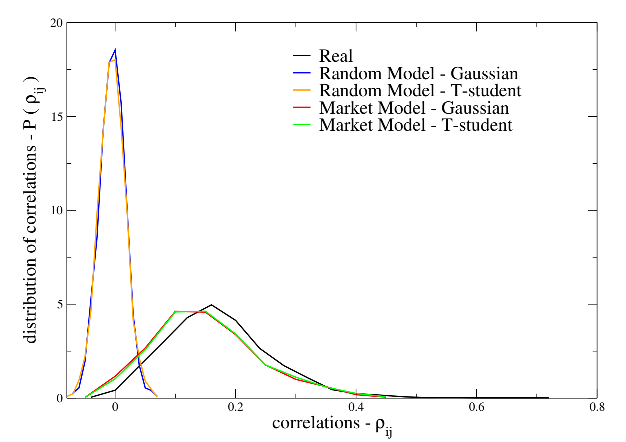
Market Model - T-student

Full time series, $T = 2322$ days (pseudo-FTSE100 stocks).

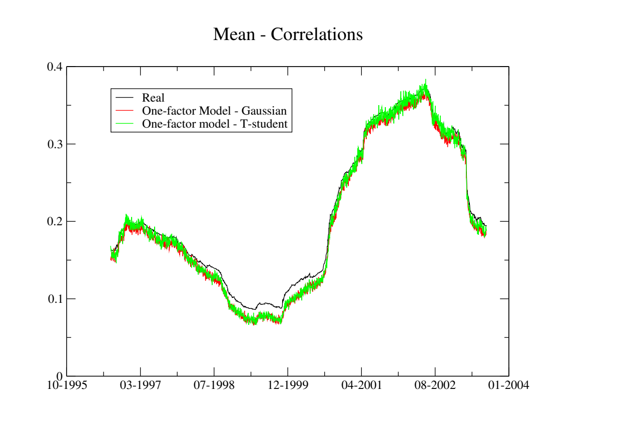


Outline

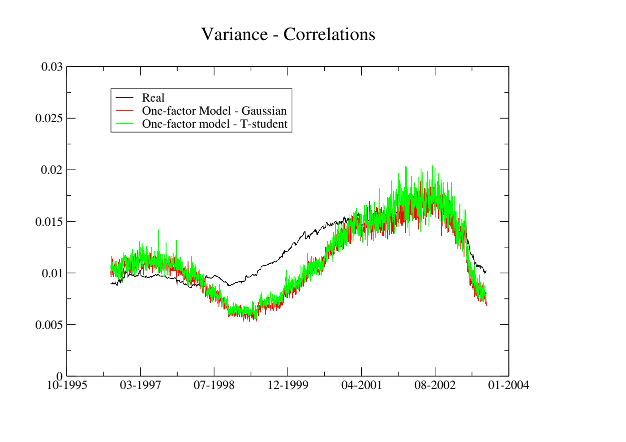
Distribution of correlations



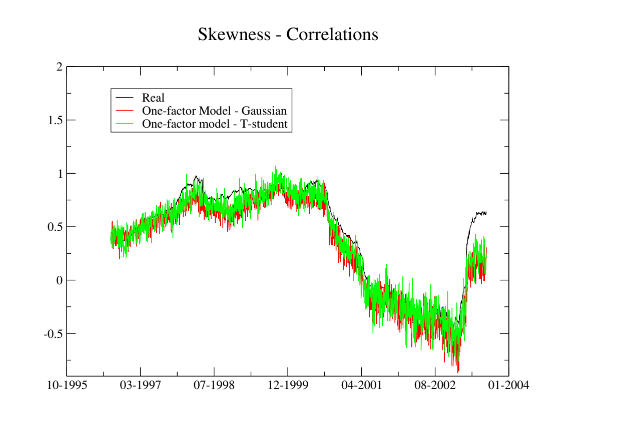
Mean correlation in time



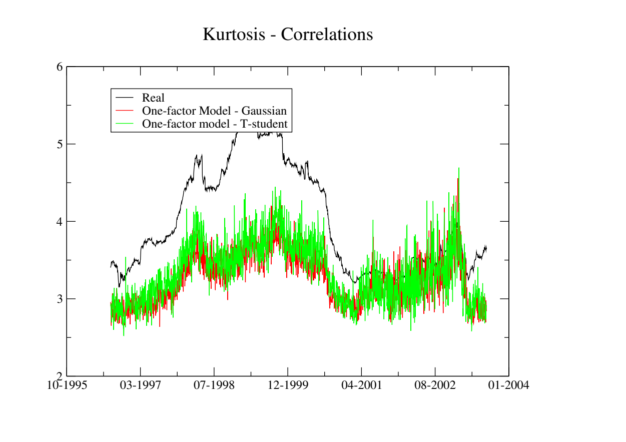
Variance of correlations in time



Skewness of correlations in time

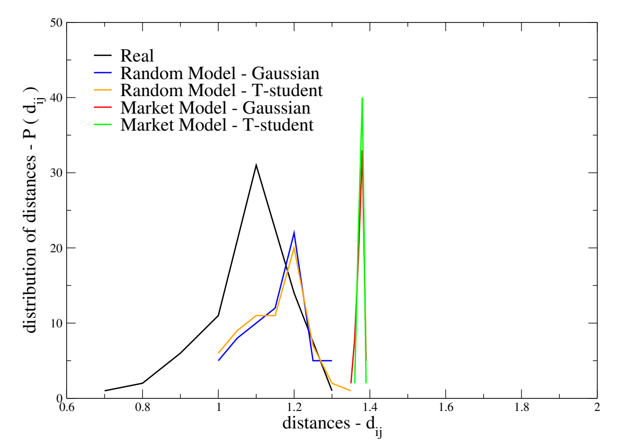


Kurtosis of correlations in time

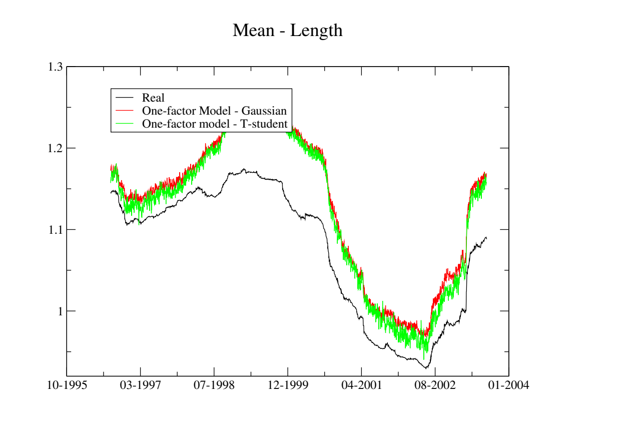


Outline

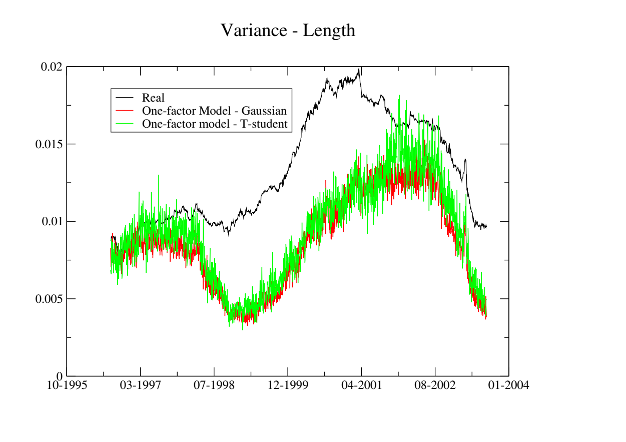
Distribution of distances



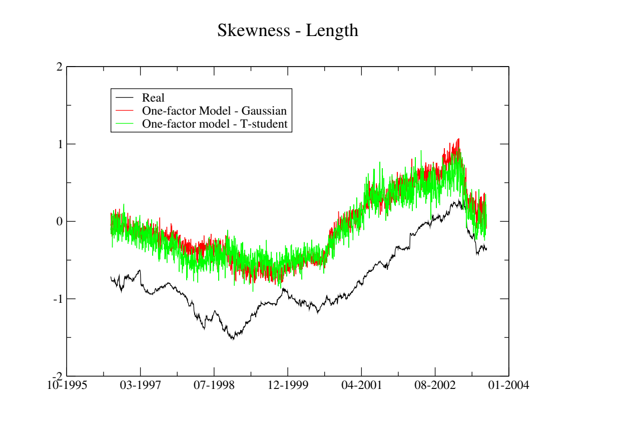
Mean distance in time



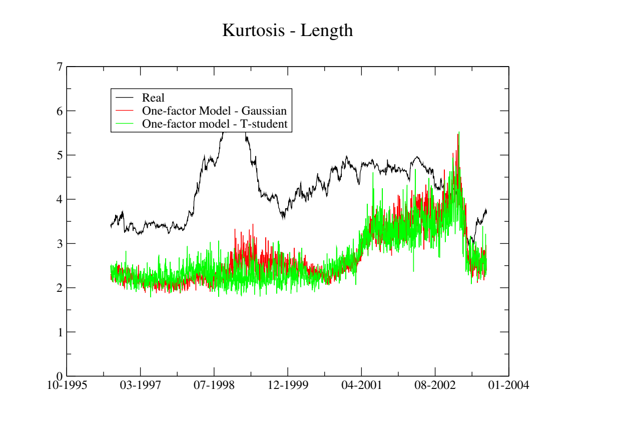
Variance of distances in time



Skewness of distances in time

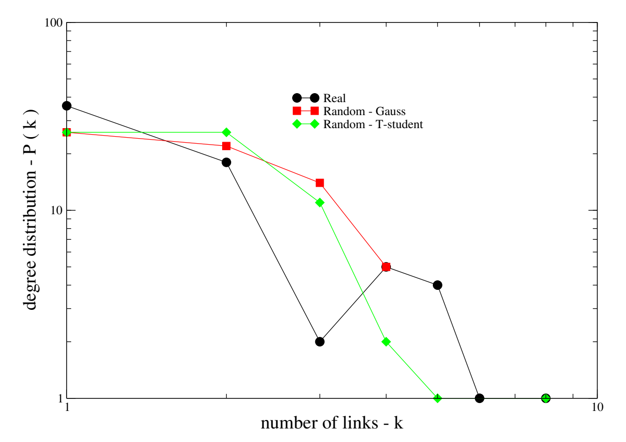


Kurtosis of distances in time

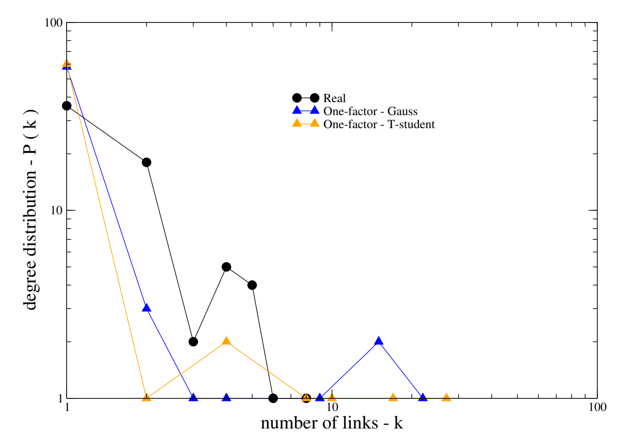


Outline

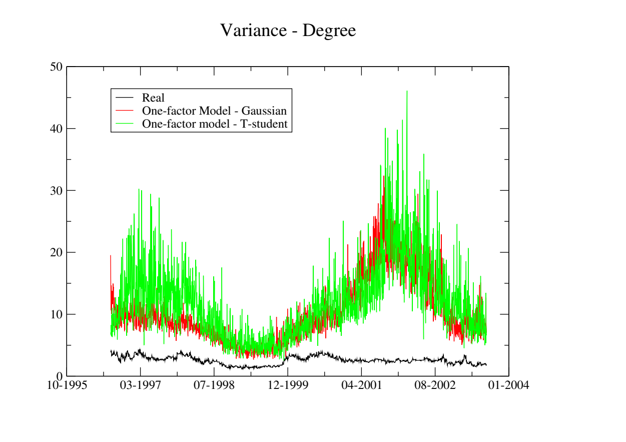
Degree distribution



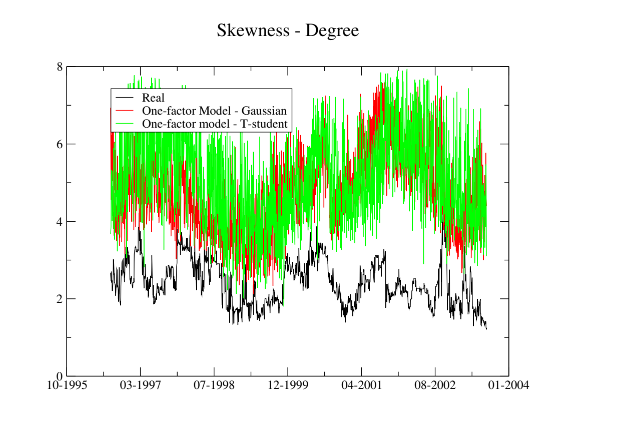
Degree distribution



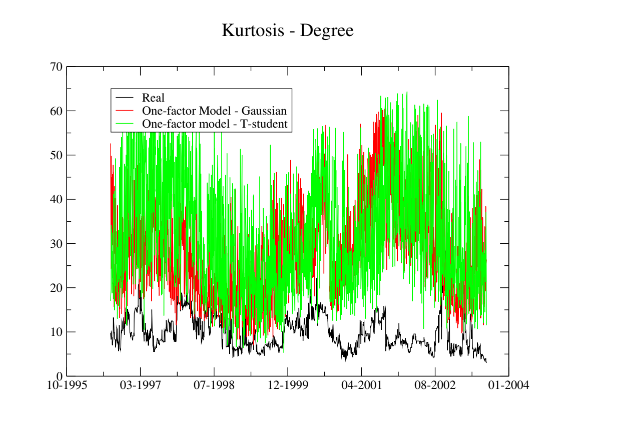
Variance of degree distribution



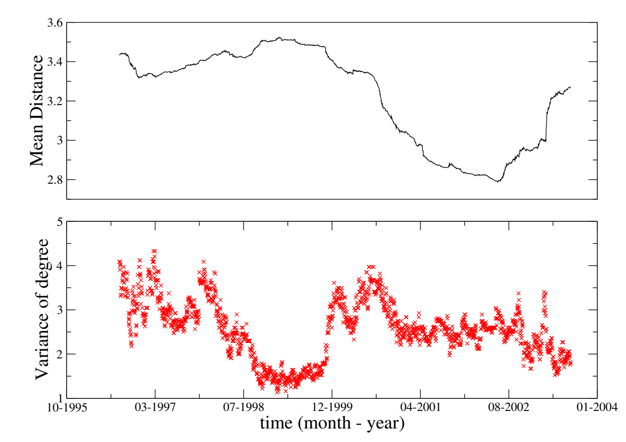
Skewness of degree distribution



Kurtosis of degree distribution



Degree versus distance



Conclusions

- Similar behaviour for both the FTSE100 stocks and the World Indices:
  - Mean and variance of correlations are correlated
  - Mean and variance of distances are anti-correlated
  - MST show different clusters (Industrial sectors / Geography)
- We can mimic the moments of correlations from random time series (*Market Model*), compared with the real one.
- The moments of distances and degree (related with the MST) cannot be mimic with these simple models.
- There is more information in these trees that we should study ...

Conclusions

- Similar behaviour for both the FTSE100 stocks and the World Indices:
  - Mean and variance of correlations are correlated
  - Mean and variance of distances are anti-correlated
  - MST show different clusters (Industrial sectors / Geography)
- We can mimic the moments of correlations from random time series (*Market Model*), compared with the real one.
- The moments of distances and degree (related with the MST) cannot be mimic with these simple models.
- There is more information in these trees that we should study ...

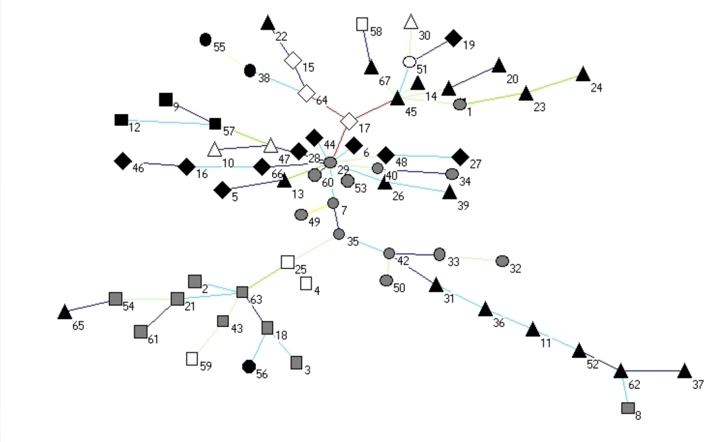
Conclusions

- Similar behaviour for both the FTSE100 stocks and the World Indices:
  - Mean and variance of correlations are correlated
  - Mean and variance of distances are anti-correlated
  - MST show different clusters (Industrial sectors / Geography)
- We can mimic the moments of correlations from random time series (*Market Model*), compared with the real one.
- The moments of distances and degree (related with the MST) cannot be mimic with these simple models.
- There is more information in these trees that we should study ...

Conclusions

- Similar behaviour for both the FTSE100 stocks and the World Indices:
  - Mean and variance of correlations are correlated
  - Mean and variance of distances are anti-correlated
  - MST show different clusters (Industrial sectors / Geography)
- We can mimic the moments of correlations from random time series (*Market Model*), compared with the real one.
- The moments of distances and degree (related with the MST) cannot be mimic with these simple models.
- There is more information in these trees that we should study ...

Future work ... (Earnings per share)



Acknowledgements

- Science Foundation Ireland - 04/BRG/PO251
- COST P10
- Shane Huston (UREKA project)