Analyzing Socio-Economic Phenomena using Physics
II. Financial Markets: Cross-correlations

Sitabhra Sinha

IMSc Chennai

Risk management

• Important element of risk management : Estimation of correlations between price movements of different stocks.

• Probability of large losses by a certain portfolio dominated by correlated moves of its constituents.

Average return $R_P = \Sigma_i \, p_i \, R_i$

p_i: amount of capital invested , R_i: expected return

Risk $\sigma_P^2 = \Sigma_{i,j} \, p_i \, C_{ij} \, p_j$

C: covariance matrix

• Optimal portfolio minimizes risk for a given value of R_P: a linear problem involving inversion of C

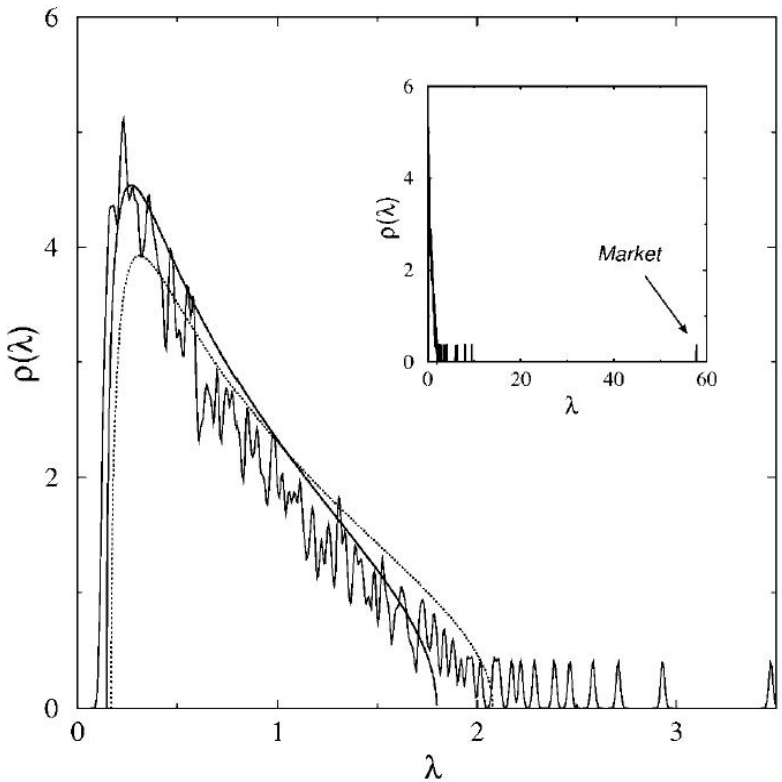
• Composition of least risky portfolio has large weight on eigenvectors of C with smallest eigenvalues

Problem: Empirical determination of C from finite time series

How random are financial correlation matrices ?

Small time series ⇒ Measurement noise

Smallest eigenvalues of correlation matrix most sensitive to this noise – the corresponding eigenvectors precisely the ones that determine the least risky portfolio!



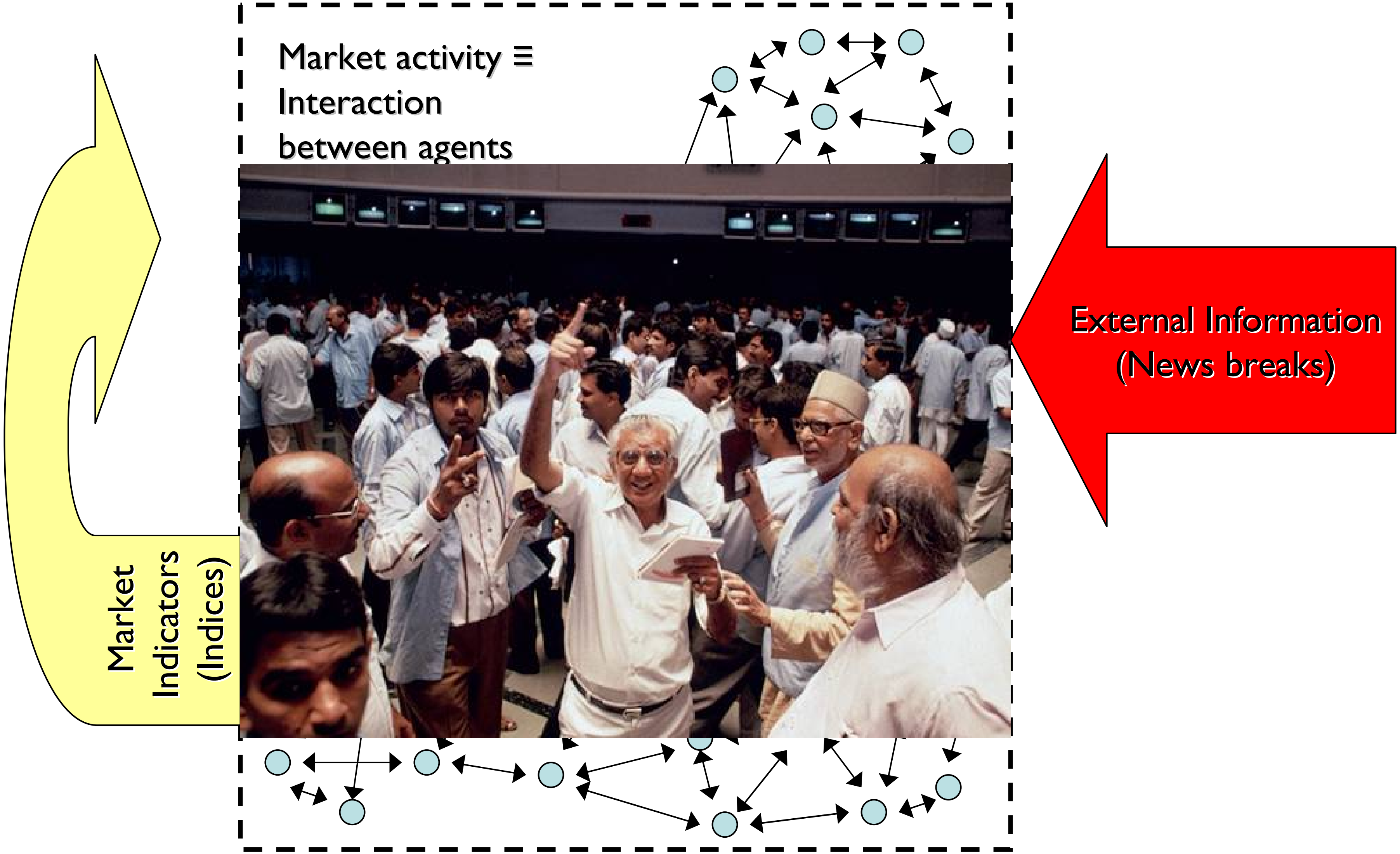
Markowitz portfolio optimization scheme based on historical determination of C not adequate since the lowest eigenvalues are dominated by noise !

Comparison of eigenvalue distribution of C with purely random matrix : Deviation only for the largest few eigenvalues!

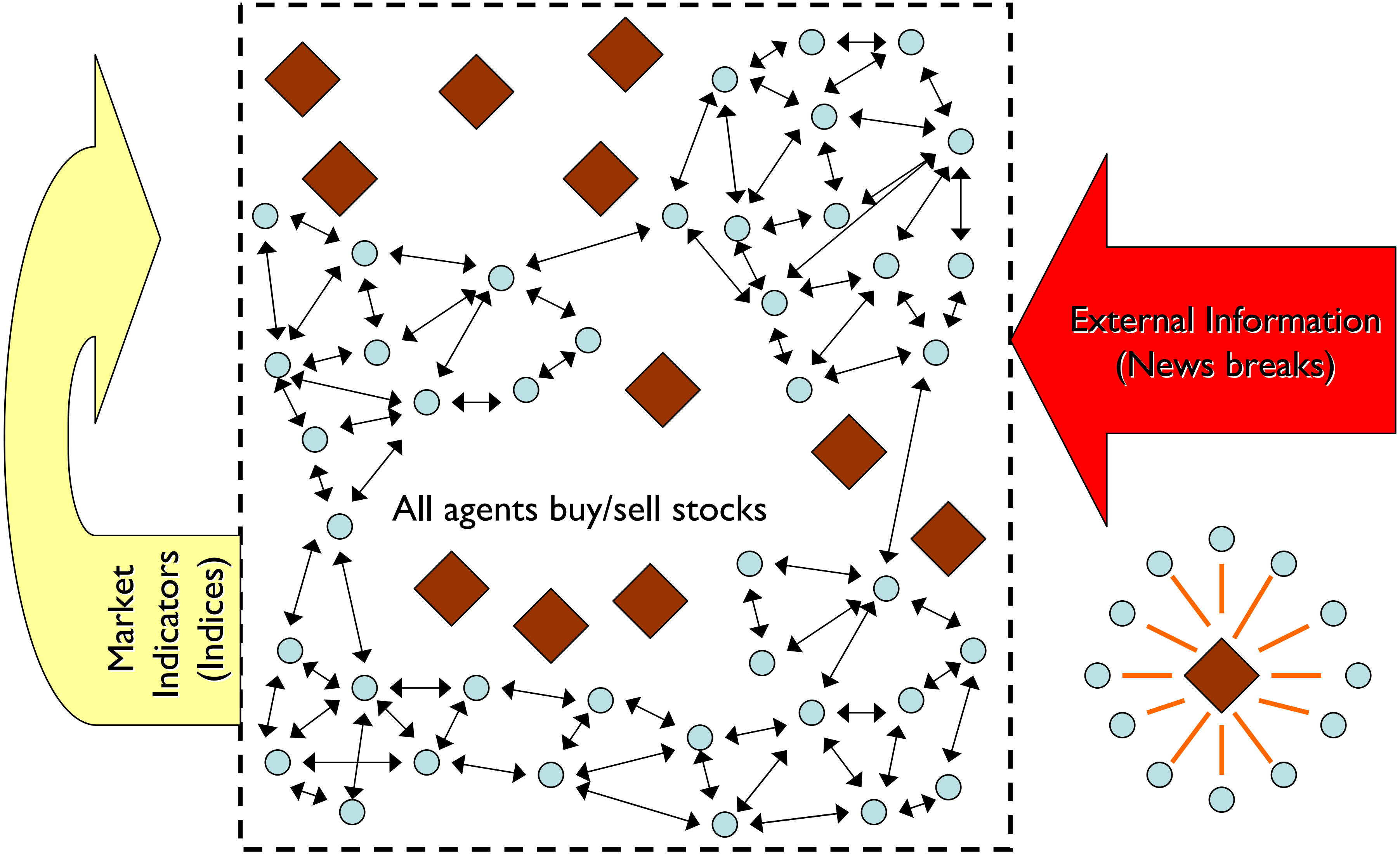
Question

How to reconstruct the interaction between different elements of the market from the data about individual price movements ?

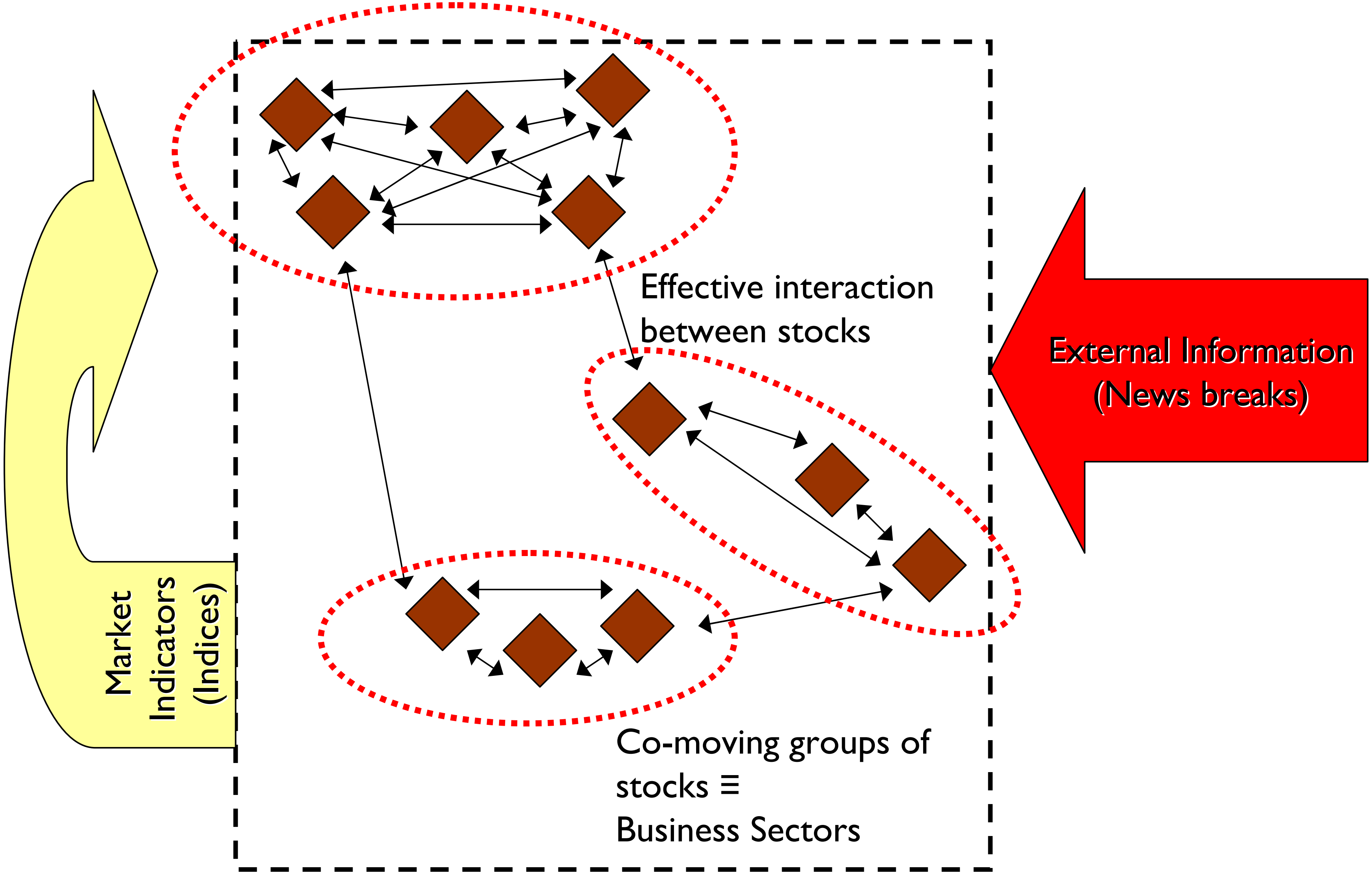
The Market ≡ Complex System



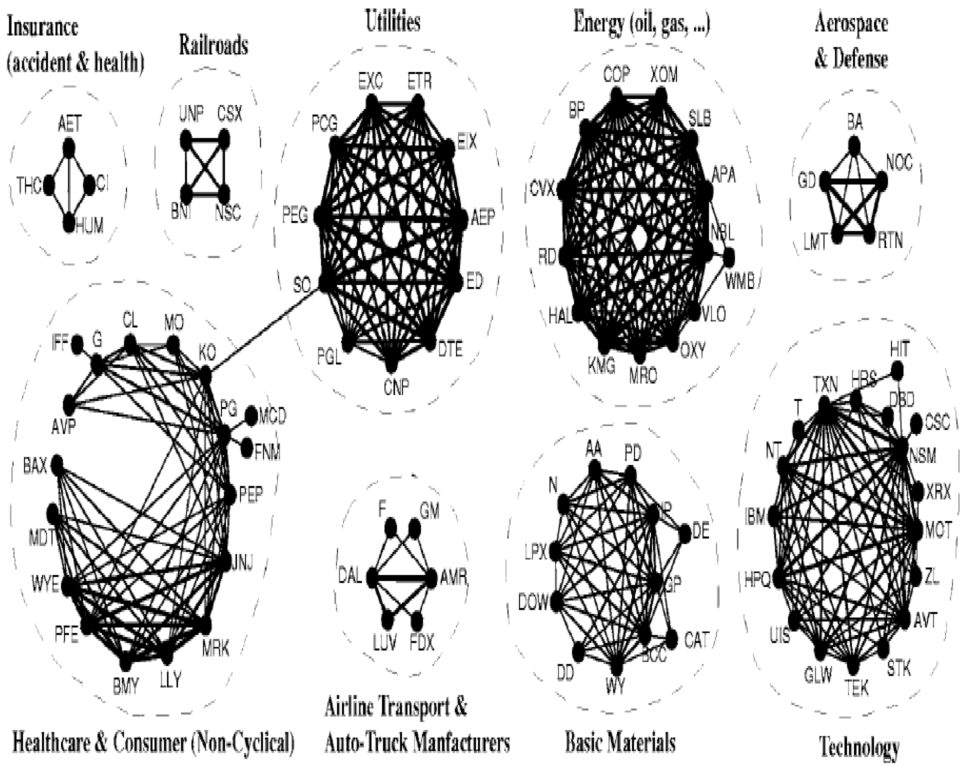
Is simpler description possible ?



"Brownian motion"



The Market seen as a System of Interacting Stocks



Market Indicators (Indices)

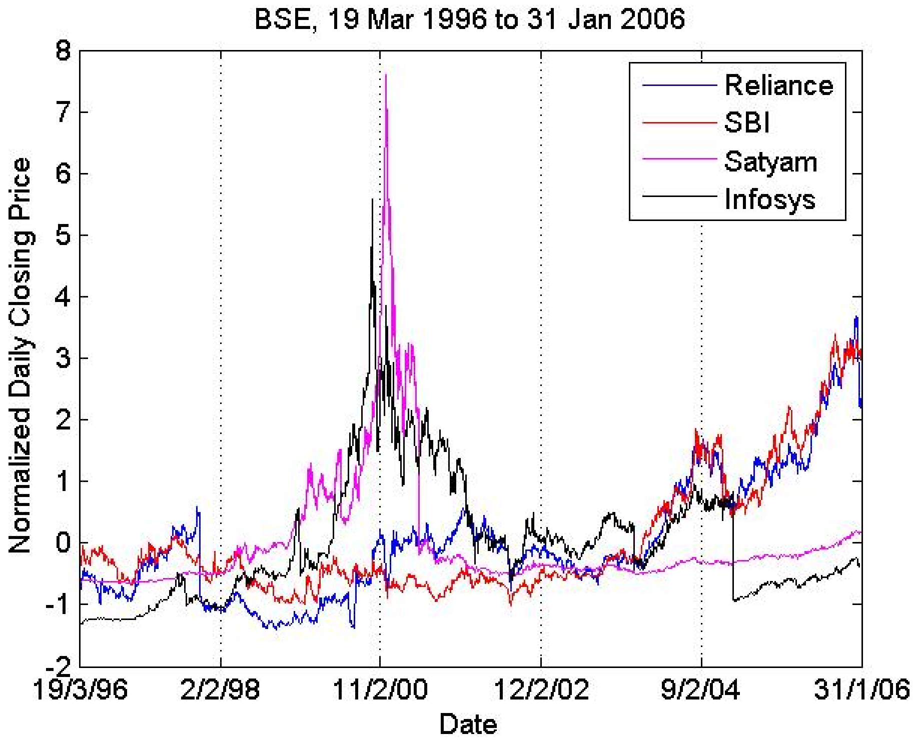
External Information (News breaks)

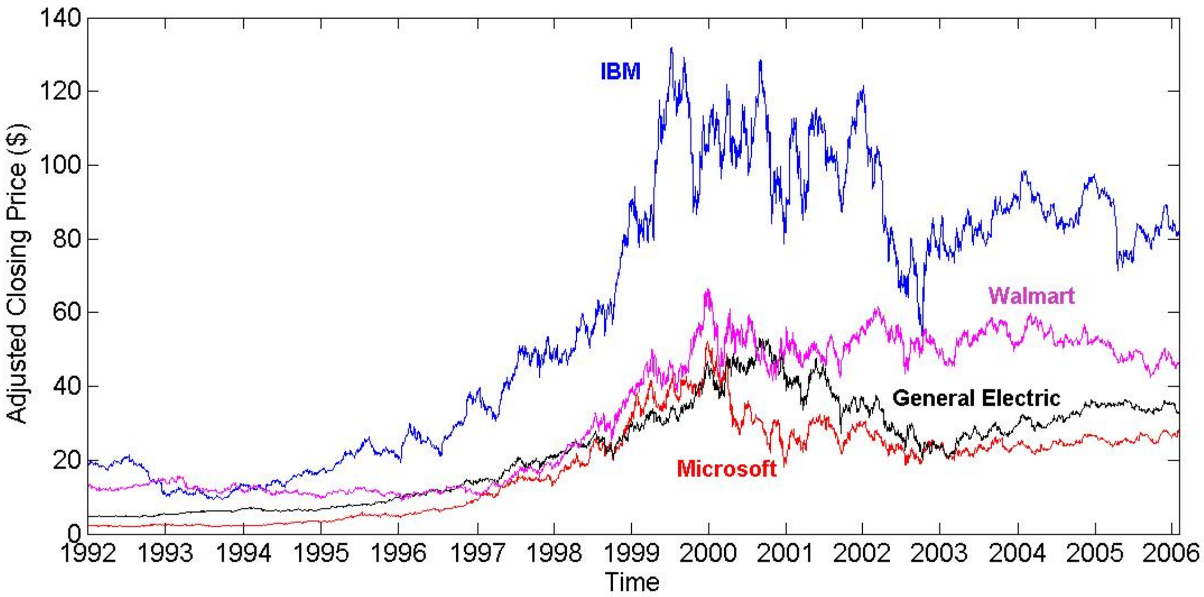
Kim & Jeong, PRE (2005)

Questions, Questions, Questions

- Can the structure of a market be empirically determined ? ⇨ Correlated movement of stocks
- But … why do emerging markets appear to be more correlated than developed ones ? ⇨ Non-interacting stocks responding to same external signals can appear *more* correlated
- What is the interaction structure of emerging markets ? ⇨ The Indian market
- What role do interactions (vis-à-vis overall market effects) play in determining observed statistical properties of markets ? ⇨ Cross-correlation of price fluctuations within/between sectors
- Can markets having similar statistical properties have very different internal structure ? ⇨ NYSE and NSE

In the Indian market
all stocks are, on the whole, remarkably correlated in their price movements !





In the NYSE
stocks movements don’t seem to be as correlated as NSE !

Question:

Can we make a quantitative comparison of the degree of correlated stock movement in the two markets ?

Answer:

Yes, by spectral analysis of the correlation matrix

Correlation Analysis

1. Construct the correlation matrix C composed of correlation values between every pair of stocks

Correlation between returns for stocks i and j:

$$C_{ij} = \langle r_i \, r_j \rangle \quad \text{where} \quad r_i = [R_i - \langle R_i \rangle]/\sigma_i$$

Data set:

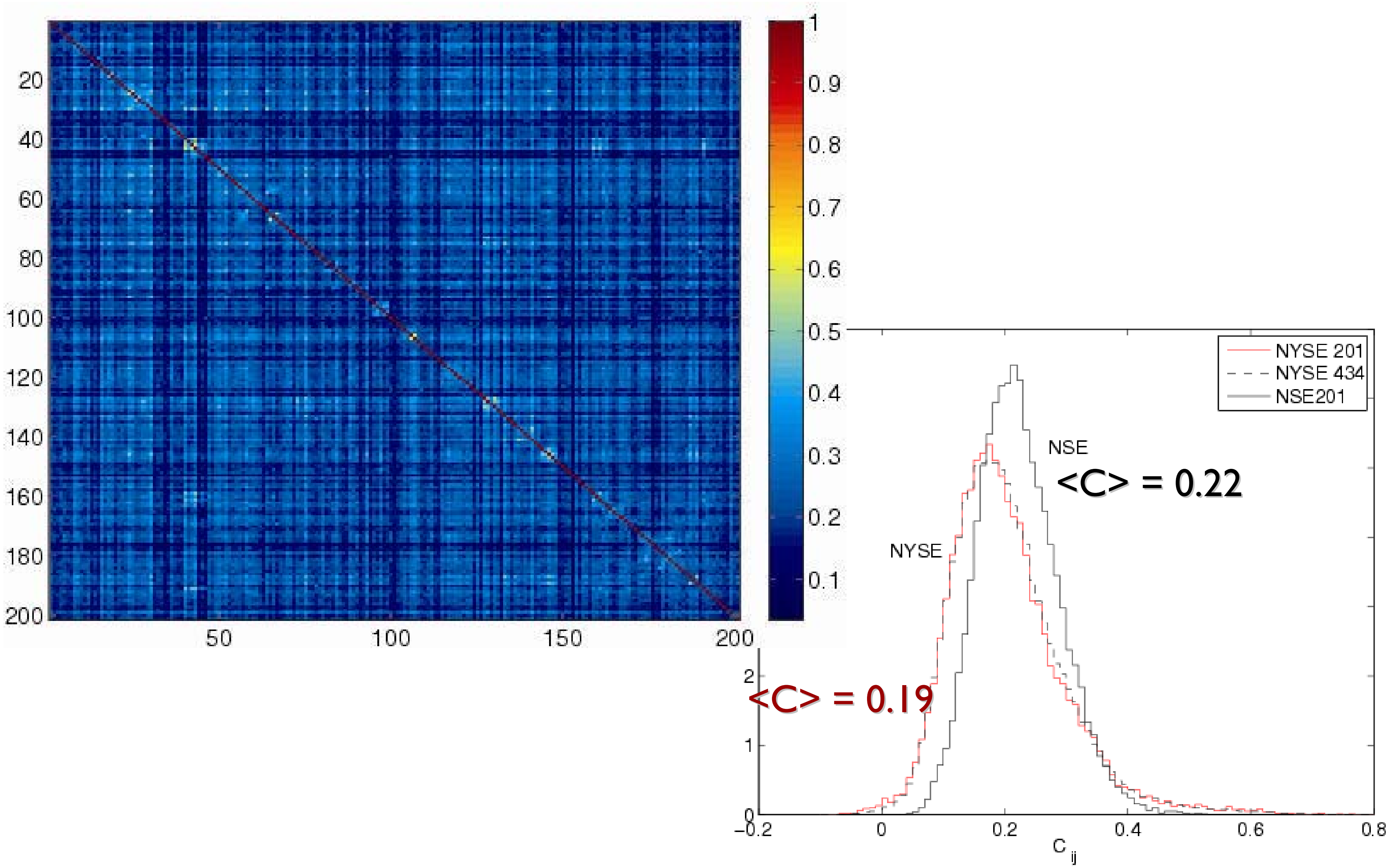
NSE: Daily closing prices of 201 stocks belonging to Nifty index from 1 Jan 1996 to 31 May 2006 (2607 working days)

Selection of stocks to minimize days of missing data

NYSE: Daily closing prices of 434 stocks [201 randomly chosen from these] belonging to S&P 500 index over same period (2622 working days)

Correlation Matrix

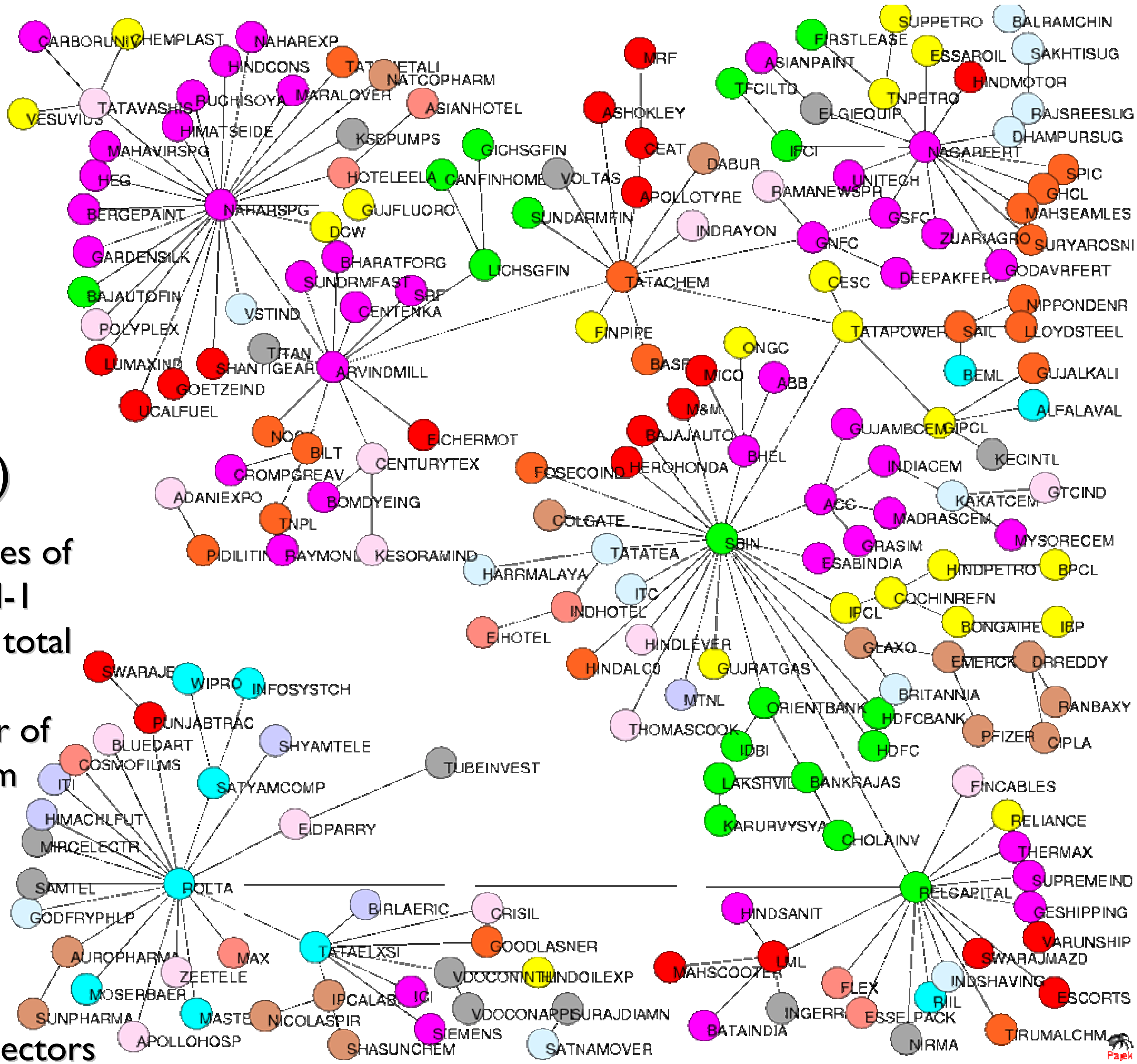
The stocks in the NSE are on average more correlated !



Minimum Spanning Tree for the Indian Market

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

Connect all N nodes of a networks with N-1 links such that the total sum of distances between every pair of nodes is a minimum



Colors indicate different market sectors

The problem with analyzing raw correlations

Cross-correlations of infinitely long time series of different variables will reflect actual inter-relations between them

But in reality, we only get data over a finite time !

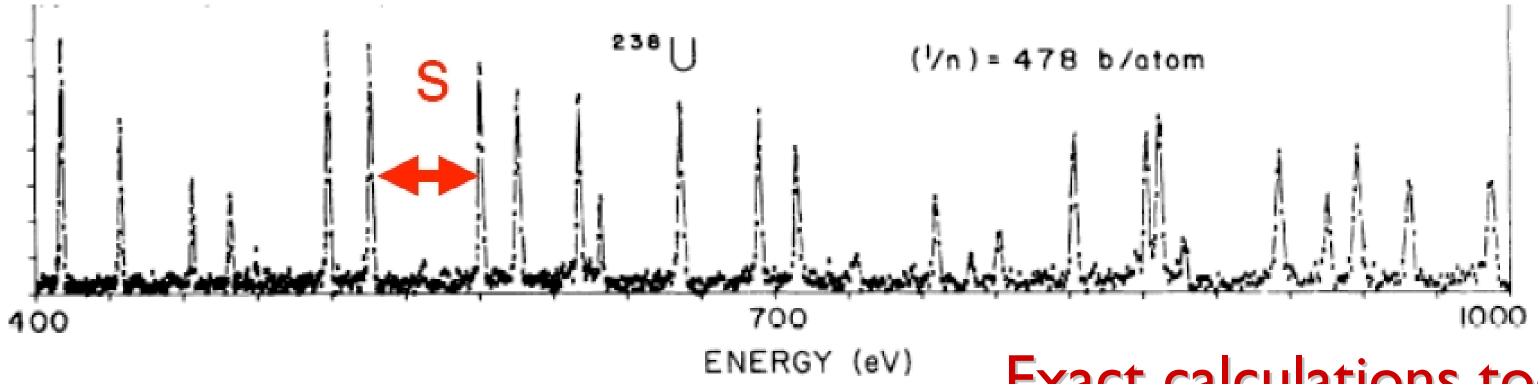
Due to stochastic fluctuations, even the output of uncorrelated random processes may exhibit spurious correlations, if we calculate C over finite-length time-series !

Question: Can we identify true interactions between variables by filtering out the effect of cross-correlations generated because of such finite time effects ?

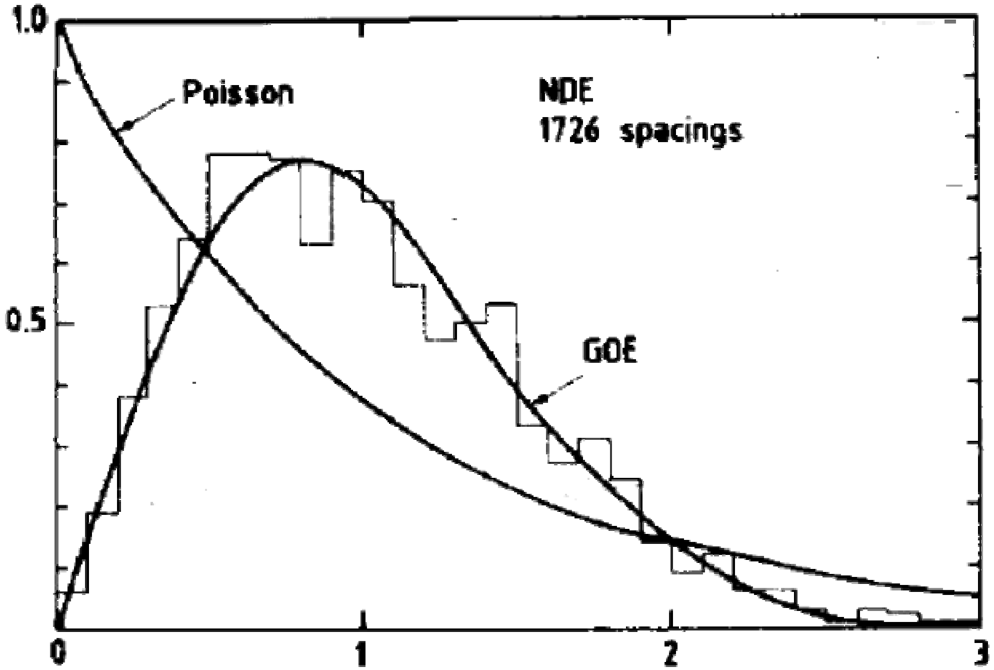
Solution: Look at the spectral properties of C and compare with that of an ensemble of random cross-correlation matrices

The spectrum of "Wigner-ium"

Nuclear physics: What is the energy spectrum of a complex nucleus ?



Exact calculations too difficult!



Wigner: Instead of focusing on specific energy levels, look at the spectral properties (eigenvalues & eigenvectors) of an ensemble of random (Hamiltonian) matrices

Distribution of spacing between neighboring energy levels

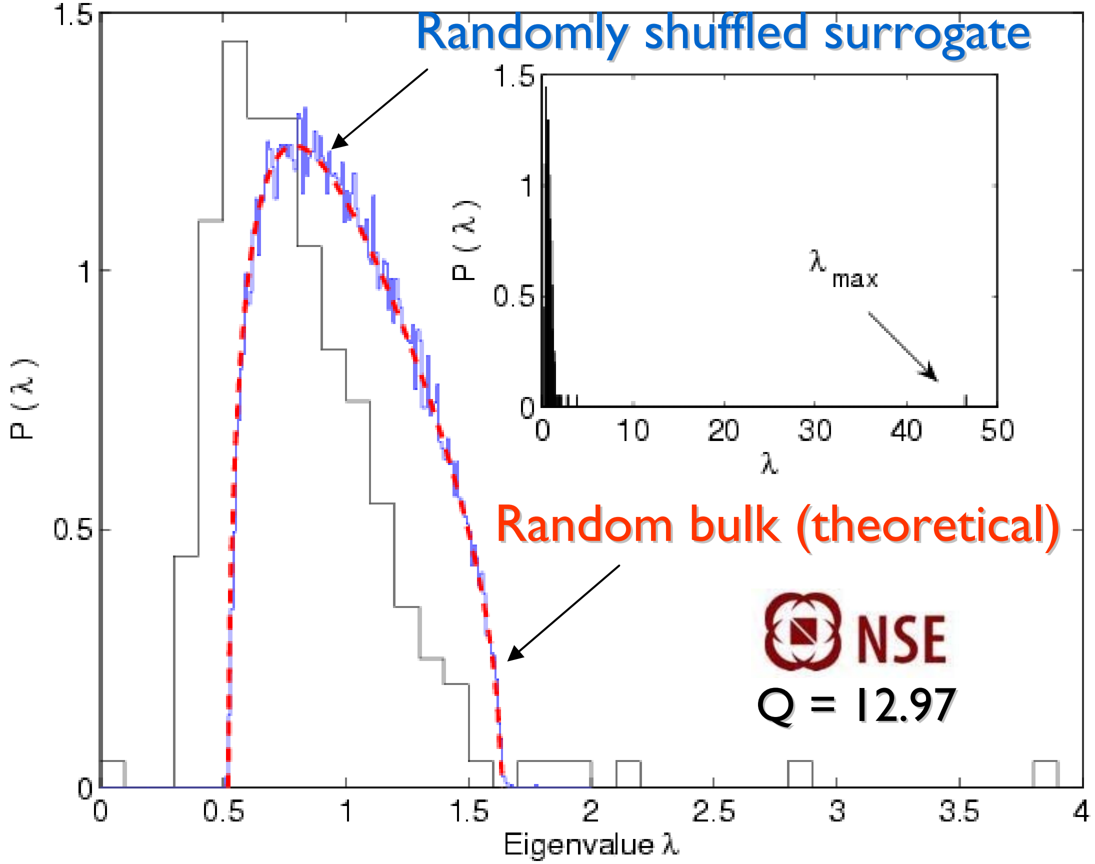
$$P(s) = s \exp(-[\pi/4]s^2)$$

II. Obtain eigenvalues of the correlation matrix

If all stocks are uncorrelated, C will be a random (Wishart) matrix with

$$P(\lambda) = [Q/2\pi] \frac{\sqrt{[(\lambda_{max} - \lambda)(\lambda - \lambda_{min})]}}{\lambda} \quad \text{where } Q = T/N$$

Bounds of random distrn : $\lambda_{max} = [1+(1/\sqrt{Q})]^2$ and $\lambda_{min} = [1-(1/\sqrt{Q})]^2$



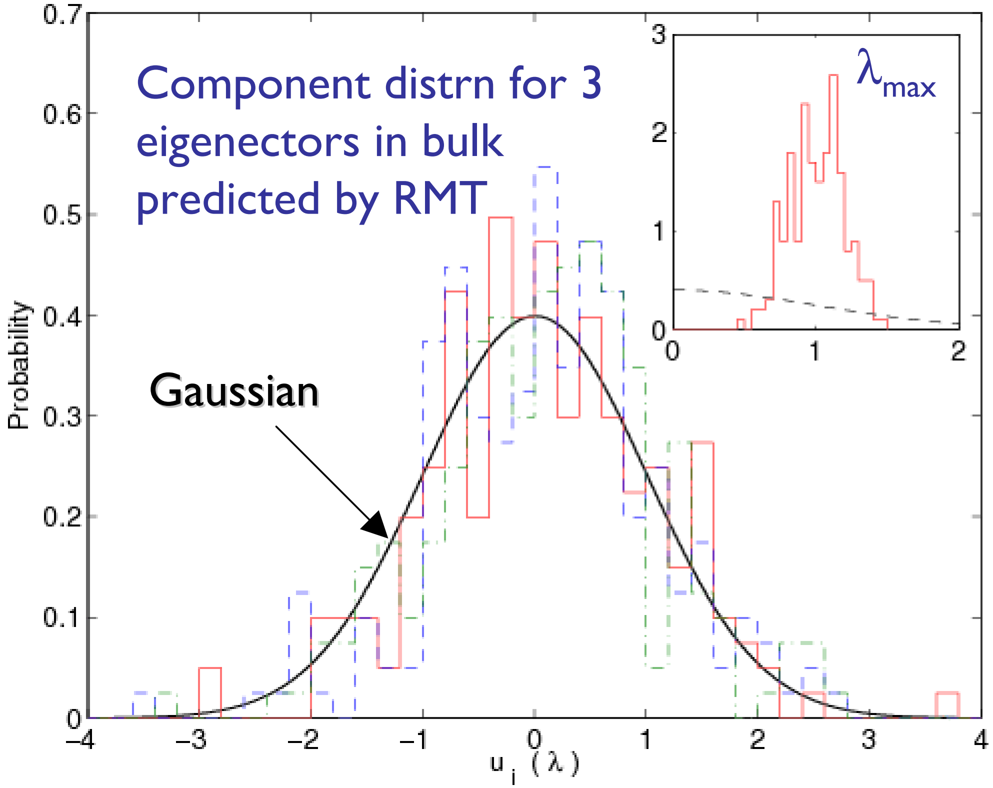
A small fraction of eigenvalues (~ 3%) deviate from random behavior

The largest eigenvalue is more than 28 times larger than the predicted max. random bound
+ a few "intermediate" eigenvalues

Random nature of the smaller eigenvalues

Let us observe the distribution of eigenvector components of the eigenvalues

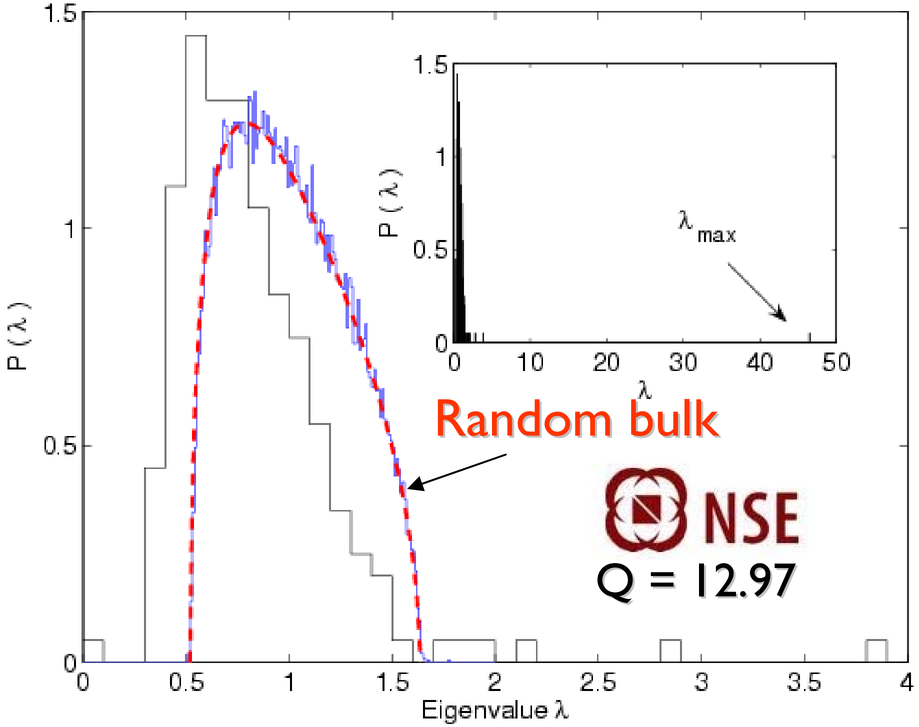
For random matrices generated from uncorrelated time-series, distrn of eigenvector components follow Porter-Thomas distrn:

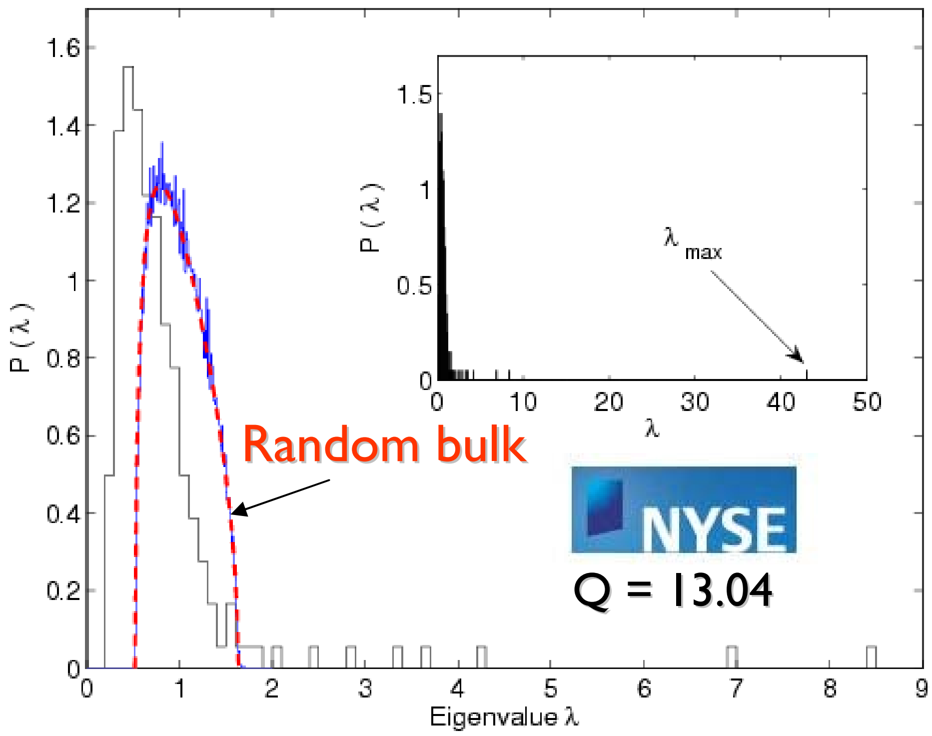


$P(u) = (1/\sqrt{2\pi})\exp(-u^2/2)$

Fits empirical data for eigenvalues belonging to bulk, but largest eigenvalues deviate significantly !

Let's compare the two markets





- The largest eigenvalue is slightly higher in NSE

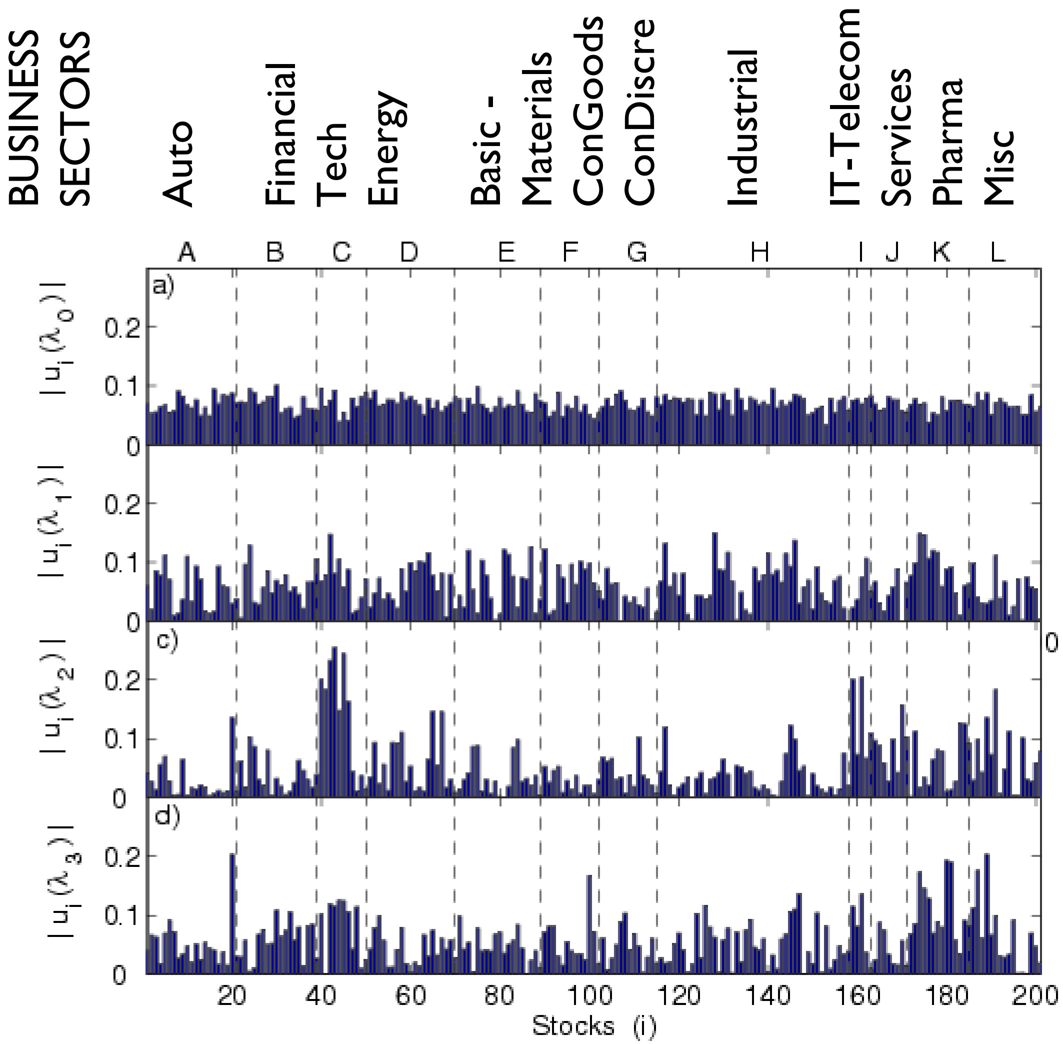
⇨ Consistent with the observation of more correlation in NSE

- The number of intermediate eigenvalues (between λ_{max} and the random bulk) is lower for NSE

⇨ Suggests relative absence of distinct structure of interactions

Deviating eigenvalues ⇨Information about interaction structure of the market

Let us look at the eigenvectors *u* of the largest few eigenvalues



All stocks contribute (almost) uniformly to *largest* eigenvalue

⇨Market mode

Common component affecting all stocks with same bias

Intermediate eigenvalues should reflect group structure in market if eigenvectors are *localized*

In NSE, no straightforward interpretation in terms of sectors

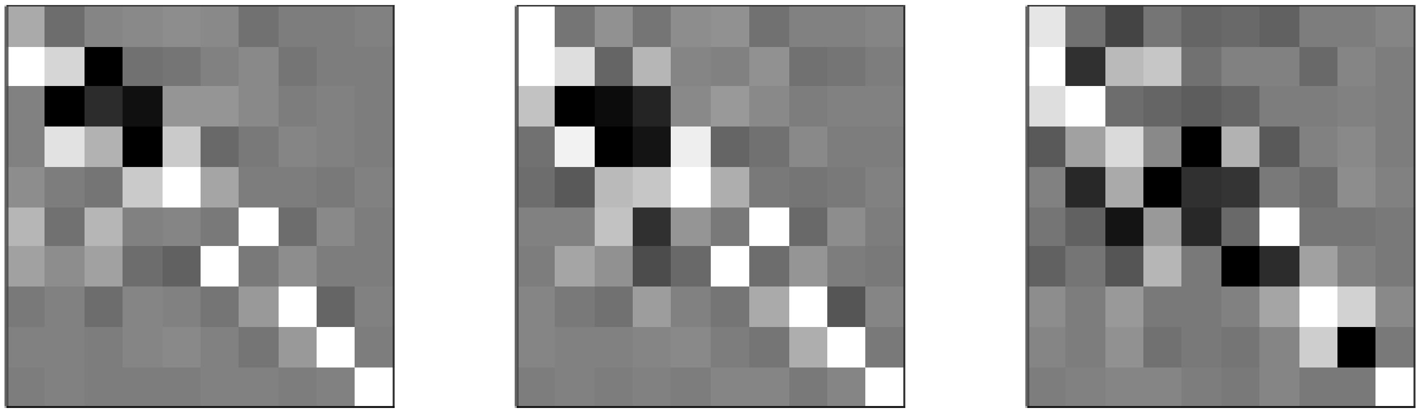
Time evolution of correlation structure

If deviations from random bulk are indicators of genuine correlations, deviating eigenvectors should be stable in time

- Consider eigenvectors of 10 largest eigenvalues of a correlation matrix D_A constructed using a temporal subset [t, t+T] of the data and D_B over a time-lagged interval $[t+\tau, t+\tau+T]$
- Calculate the overlap matrix $O(t,\tau) = D_A D_B^T$
- In ideal case (non-random eigenvectors stable in time), O = identity matrix

Overlap matrices with T = 5 years

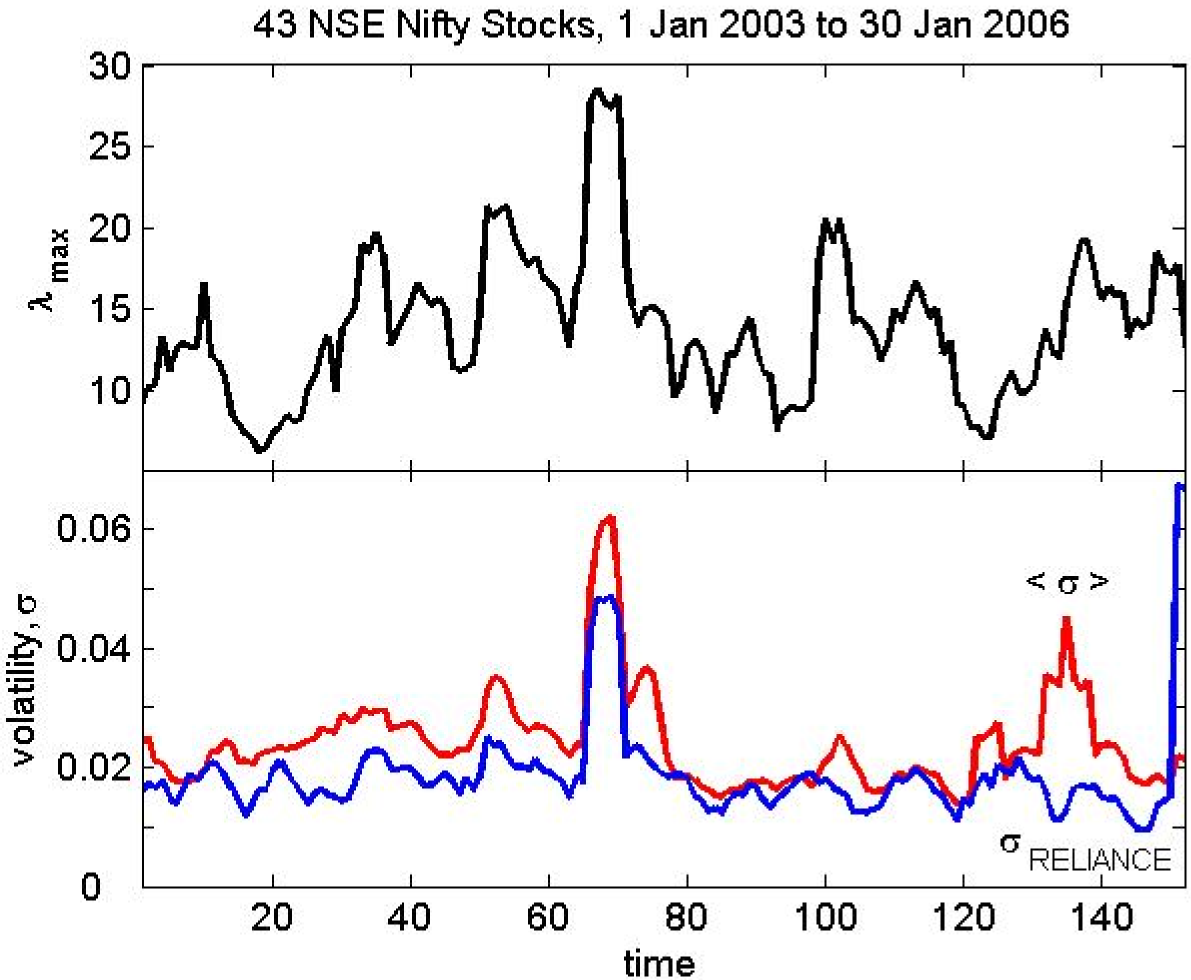
Reference period: Jan 1996



τ=6 months τ=1 year τ=2 years

Eigenvectors show different degrees of stability, with the one corresponding to the largest eigenvalue being most stable

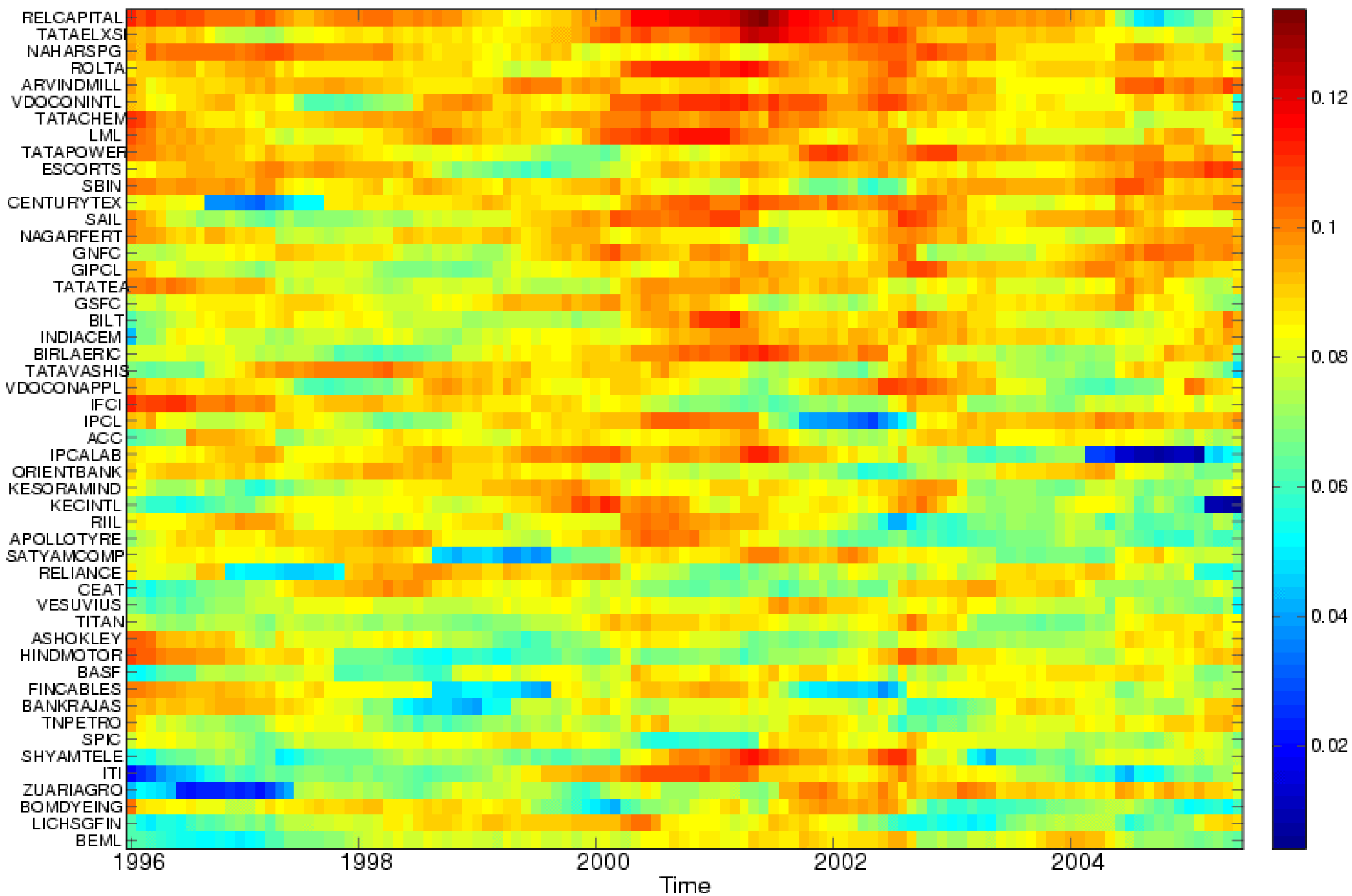
Time evolution of the maximal eigenvalue of correlation matrix : A window into volatility



Temporal evolution of composition of eigenvector corresponding to the largest eigenvalue

Allows us to identify the prime movers & shakers of the NSE !

Analysis: divide time series into overlapping windows of length T = 6 months



50 stocks having biggest contribution to eigenvector corresponding to largest eigenvalue over the period Jan'96-May'06

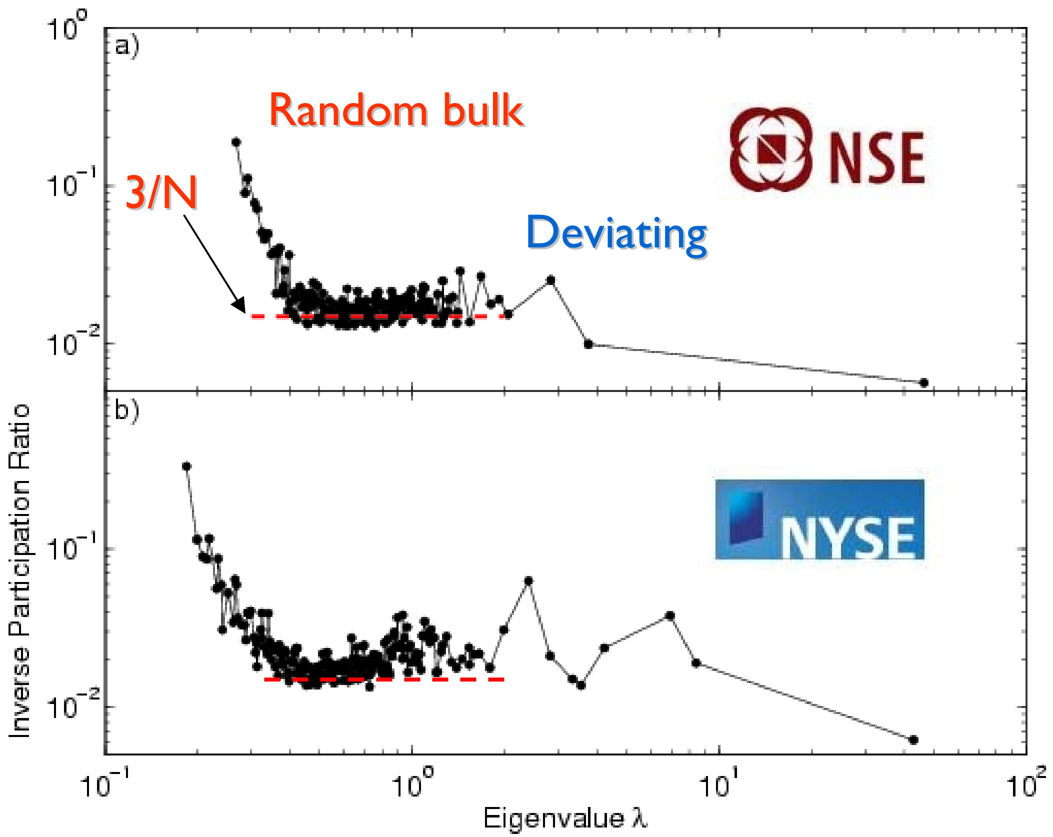
Note 5 Tatagroup companies & 3 Reliance group companies

Analyse the composition of eigenvectors for detecting localization: we look at

Inverse Participation Ratio

For the jth eigenvector $I_j = \Sigma_{i=1,\ldots,N} [u_{ij}]^4$, u_i: eigvector components

• If all components are equal ($u_i = 1/\sqrt{N}$), as for λ_0 ⇨ $I=1/N$
• Dominant contribution of a single component ($u_1=1$ & $u_i=0$, $i=2,\ldots,N$) ⇨ $I = 1$



• I inversely related to number of significantly contributing eigvector components [stocks]

•For eigenvectors of rand corr matrix ⇨ I=3/N [seen in the random bulk]

• I > 3/N for most deviating eigenvectors ⇨Localization

• But much less significant in NSE compared to NYSE

However …

largest eigenmode (*market*) dominates all intra-group correlations (if existing).

$\Rightarrow$ no straightforward detection of significantly related groups of stocks.

For this purpose, use

Matrix Decomposition Technique

Aim: removing the effect of (i) market mode & (ii) random noise

Expanding correlation matrix as $C = \Sigma_i \lambda_i u_i^T u_i$

Allows decomposition of C into contributions due to

- *market*, common for all stocks
- *groups* of co-moving stocks (identified with various business sectors)
- *random*, idiosyncratic effects for each stock

$$C = C_{market} + C_{sector} + C_{random}$$

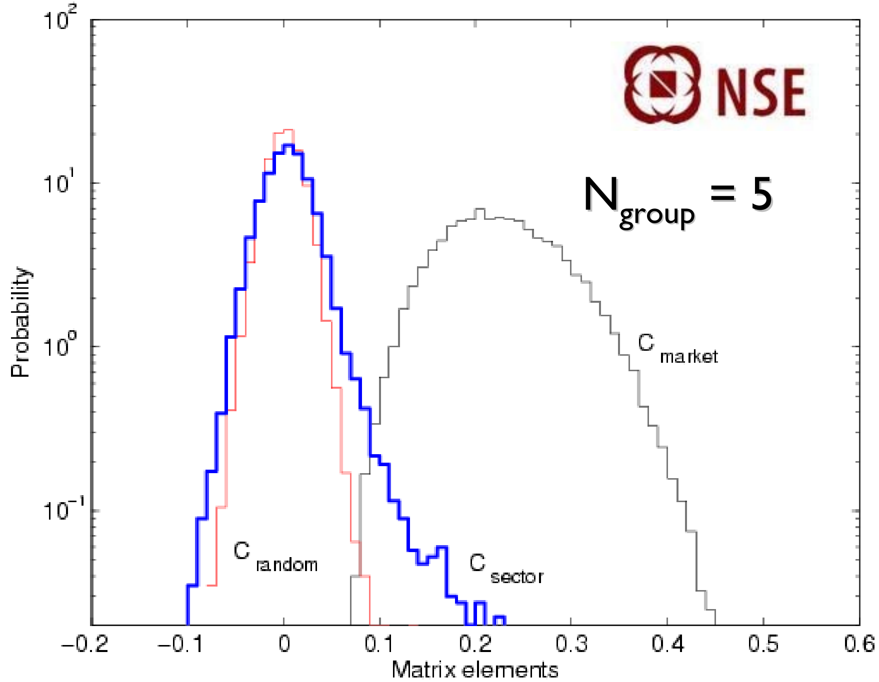
$$= \lambda_0 u_0^T u_0 + \Sigma_{i=1,\ldots,N_{group}} \lambda_i u_i^T u_i + \Sigma_{i=N_{group}+1,\ldots,N-1} \lambda_i u_i^T u_i$$

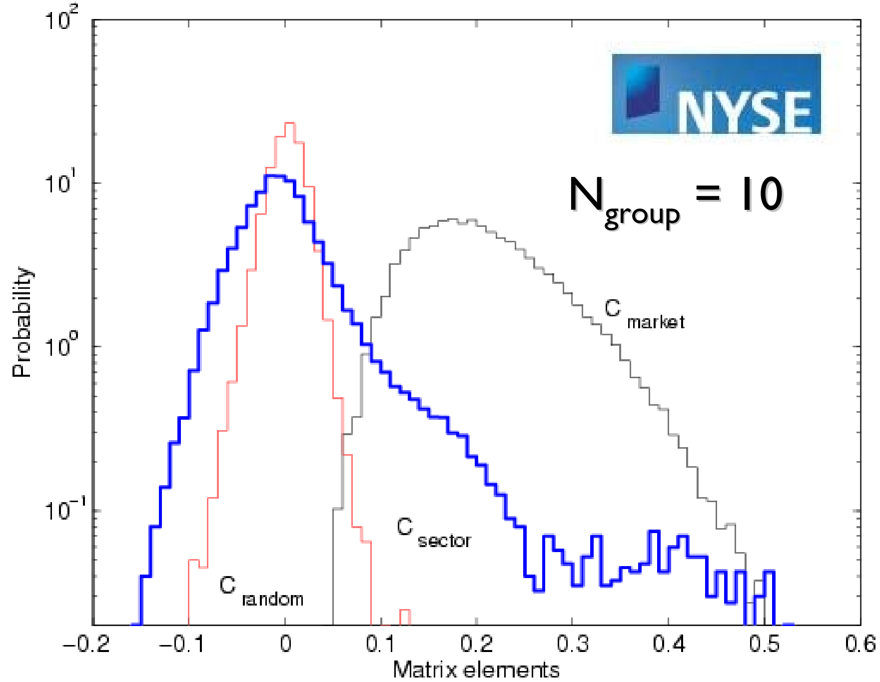
Largest eigenvalue | Intermediate eigenvalues | Random bulk eigenvalues

Filtered correlation matrices

By visual inspection of intermediate eigenvalues for NSE, we choose $N_{group} = 5$

No signficant error from choice ➪variations of N_{group} due to eigenvalues closest to random bulk ➪ has lowest contribution to C_{sector}





NYSE: sector correlations as strong as market effects

⇒ Distinct groups of interacting stocks (business sectors)

⇒ No such case in NSE, sector contribution weak

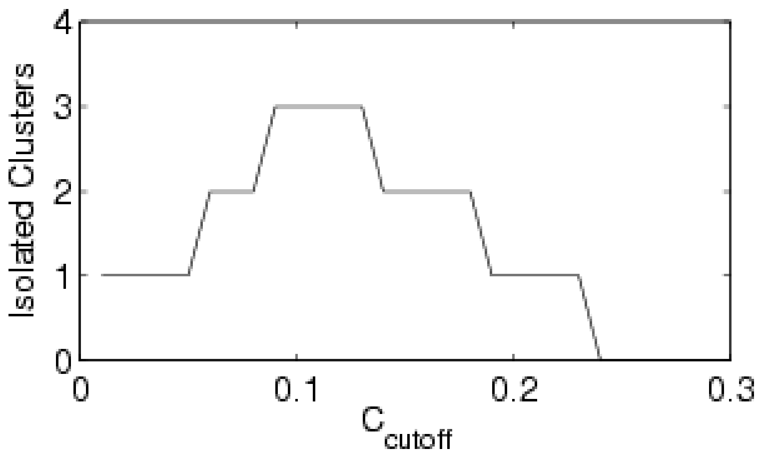
⇨NSE has much smaller fraction of significantly interacting stocks
⇨ Hard to segregate into groups having distinct sector identity

But can we show this explicitly ?

Reconstructing the stock interaction network

Method: Use the sector correlation matrix C_{sector} to generate an adjacency matrix A, such that

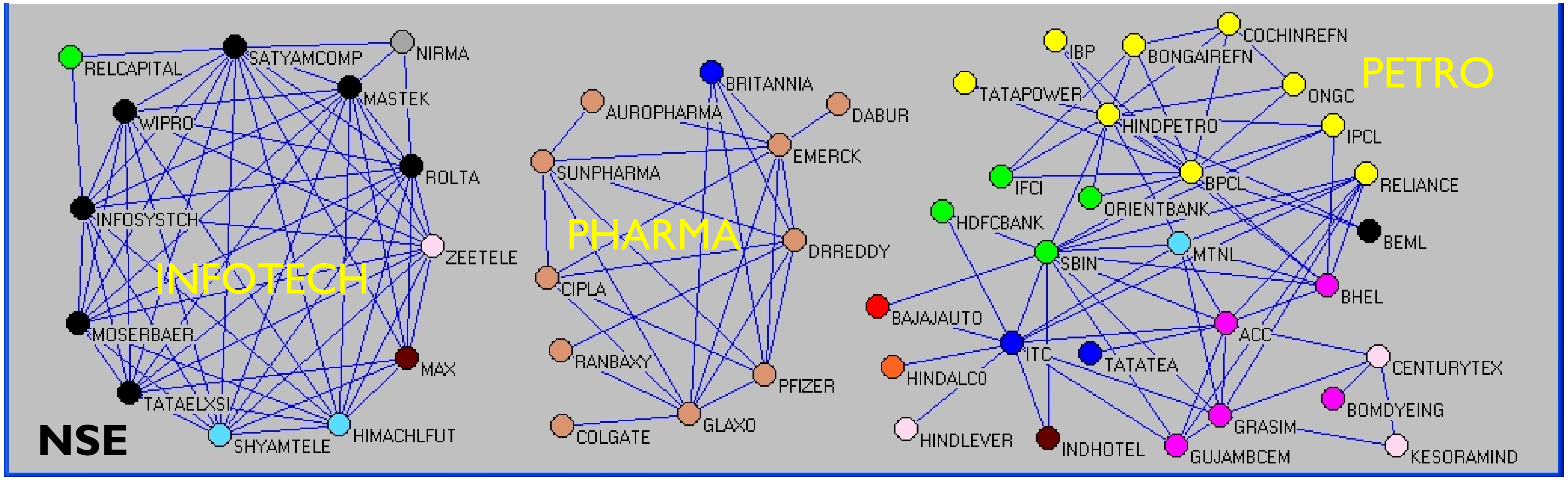
- $A_{ij} = 1$ if $C^{ij}_{sector} > C_{cutoff}$
- $A_{ij} = 0$ otherwise

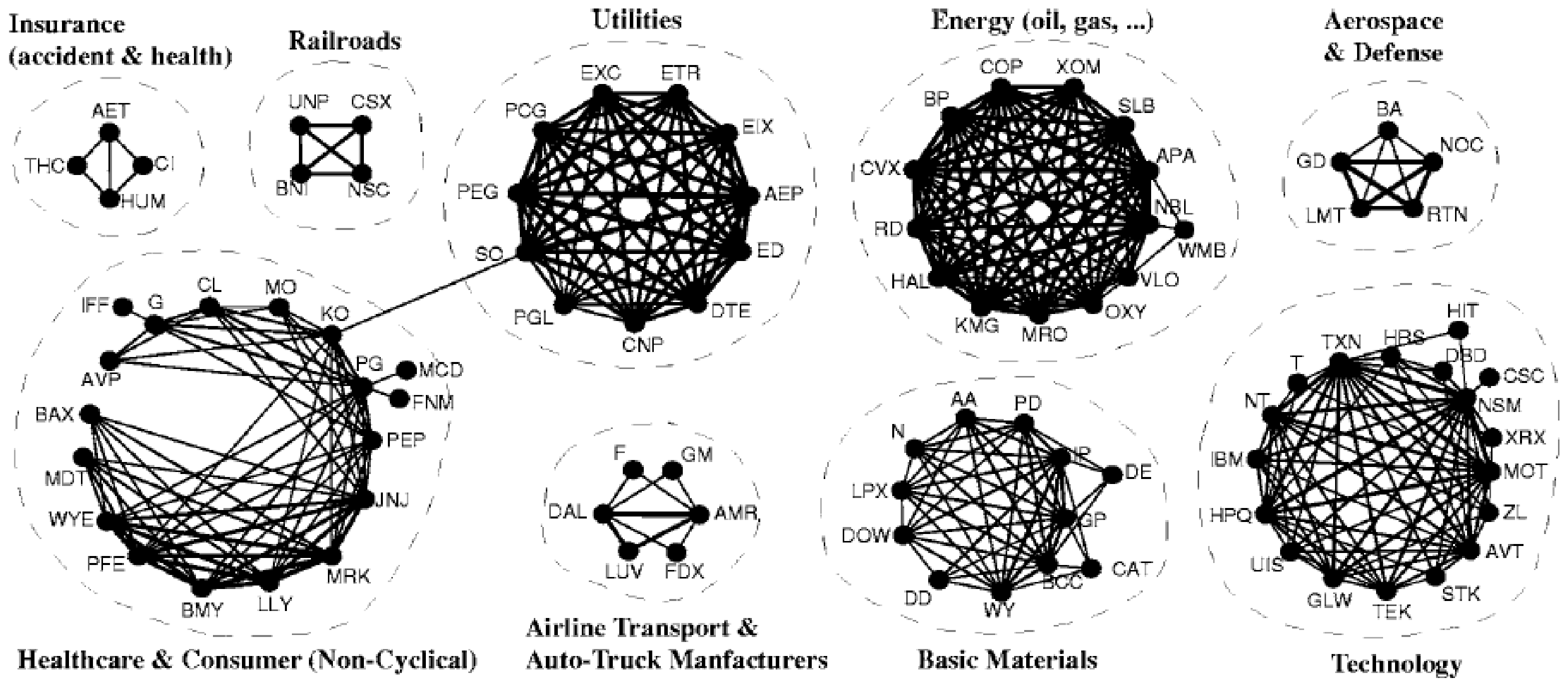


Depending on choice of C_{cutoff} we get different numbers of clusters

- C_{cutoff} small ⇨ all stocks are connected in a giant cluster
- C_{cutoff} large ⇨all stocks are isolated nodes
- Largest number of distinct clusters (=3) generated for $C_{cutoff} \sim 0.1$

Comparing Market Structures





NYSE

So… very few distinct groups of co-moving stocks can be identified in NSE compared to NSE

- But are these results too specific (i.e., valid only for NSE) ?
- Does the lack of group structure always get reflected in the eigenvalue distrn of the market correlation matrix ?
- Let's generate a random ensemble of returns with market internal structure that is specified by us & examine their spectral properties

For this, we use

Factor Models

Gaussian (0,1)

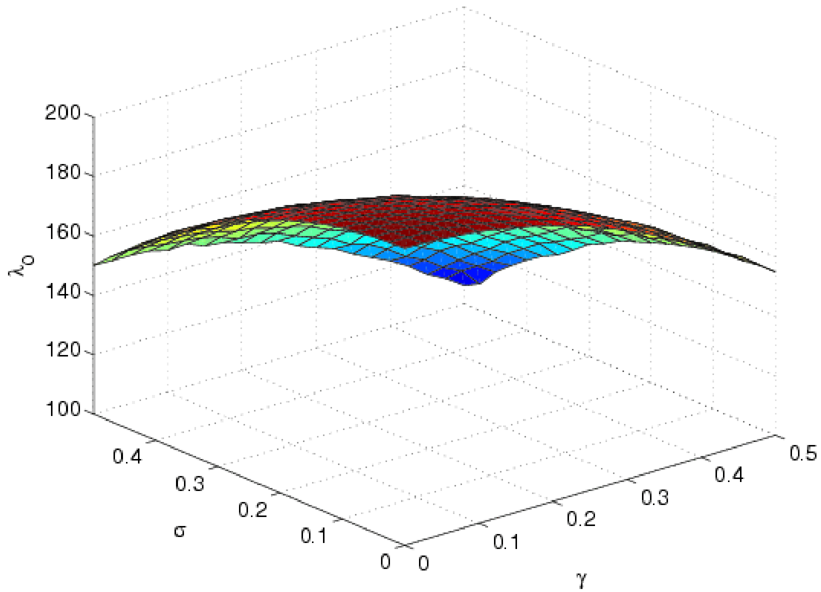
Return of i^{th} stock, $r_i^k(t) = \beta_i \, m(t) + \gamma_i^k \, \xi^k(t) + \sigma_i \, \eta_i(t)$

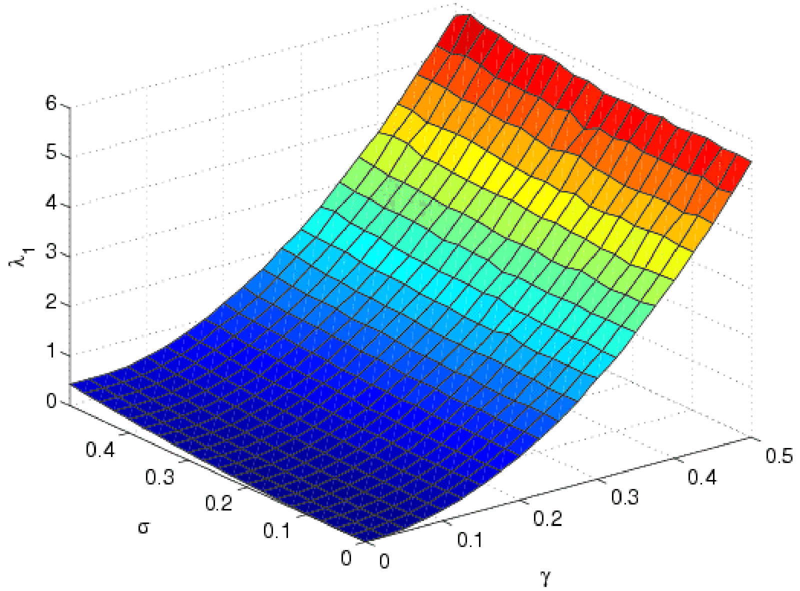
Market Sector Random

subject to the condition : $\beta_i^2 + (\gamma_i^k)^2 + \sigma_i^2 = 1$

Independent model parameters: relative strengths of sector effect (γ) & random effect (σ), and the number of sectors (M)

Simulation results with N = 200, M = 10





$\lambda_0 = N\,\beta^{\,2}$, $\quad \lambda_1 = S_{max}\,(1 - \beta)^{\,2} = S_{max}\,[1-\sqrt{(1-\gamma^{\,2}-\sigma^{\,2})}]^2$

where S_{max} : size of largest sector (=20)

- Increase of largest eigenvalue with increasing market effect
- Decrease of 2nd largest eigenvalue with dominance of market effect

$\Rightarrow$ intermediate eigenvalues occur closer to random bulk (since Q=constant)

- If sectors are of same size, intermediate eigenvalues cluster together
  - for different sized sectors, they are spaced apart
- Consistent with empirical NSE data; holds for γ & σ distributed over a range

Conclusions

❑ A detailed study of internal structure (network of stock interactions) of an emerging market

• Supports the notion that emerging markets are more correlated than developing ones ⇨ implication for portfolio diversification (risk reduction) for investing in such markets

❑ Reason for correlation:

• Lack of distinct sector identity, relative absence of groups of co-moving stocks

• Market effects dominate, systems tends to move as a homogeneous entity to information shocks (e.g., news breaks)

• Hypothesis: Gradual emergence of sectors as market matures

❑ Implication for observed statistical properties of markets

• Universality of “inverse cubic” law $\Rightarrow$ Price fluctuation distrn may be explained as the response behavior of a *single* entity

(no need to consider complex internal structure of the market)