Stochastic Systems Group

Exploiting Sparse Markov *and* Covariance Structure in Multiresolution Models

Myung Jin Choi

Joint work with Venkat Chandrasekaran and Alan Willsky

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

June 15, 2009

Stochastic Systems Group

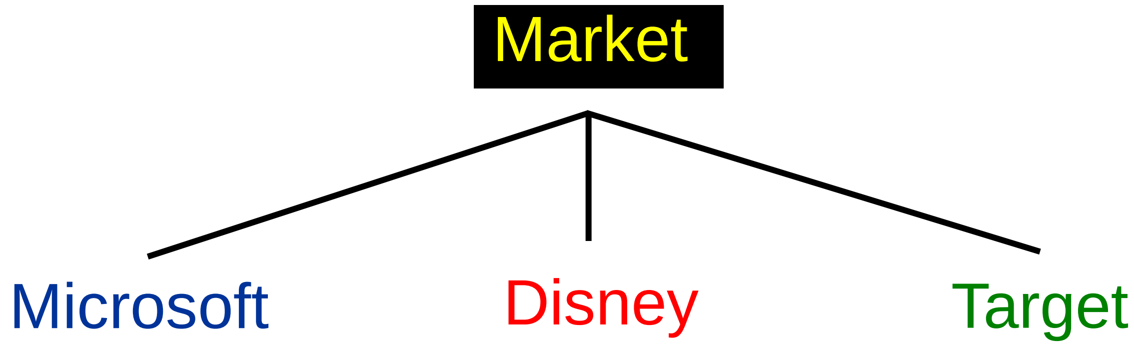
Dependency Structure of Monthly Stock Returns

- The monthly return of one stock is correlated to all other stocks.

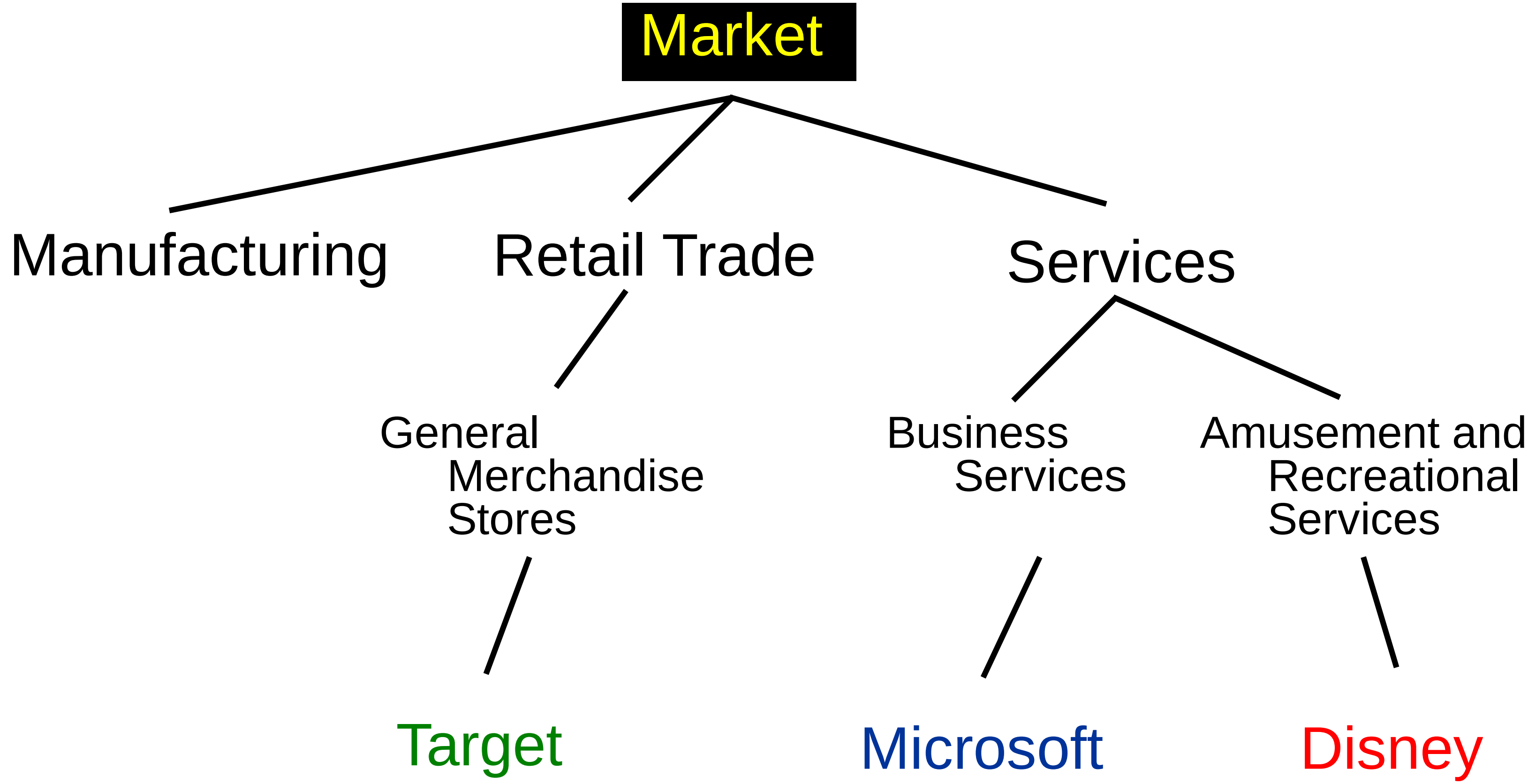
Dependency Structure of Monthly Stock Returns

- Because stock prices tend to move together driven by the market situation.

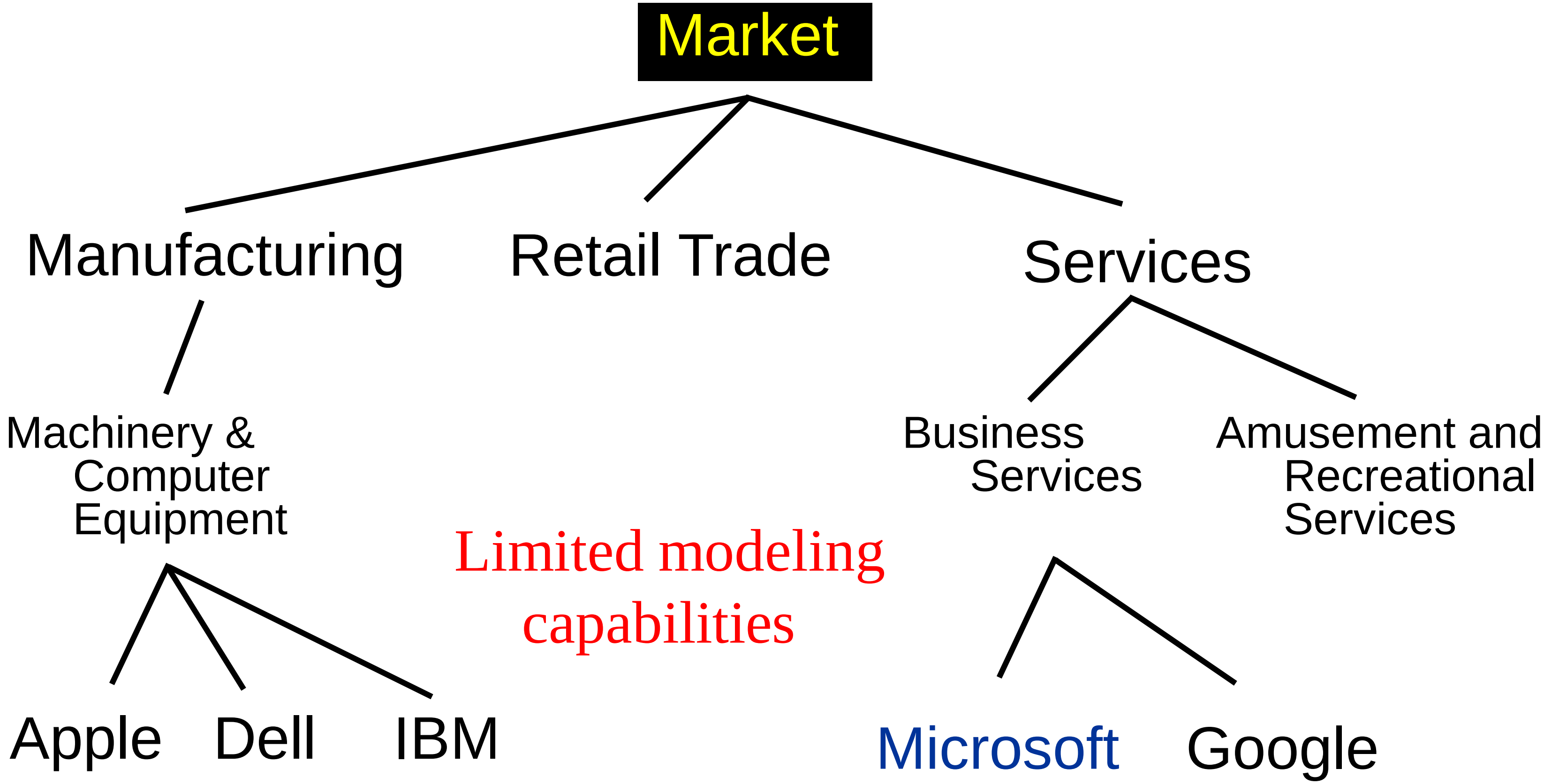
Multiresolution (MR) Tree Models



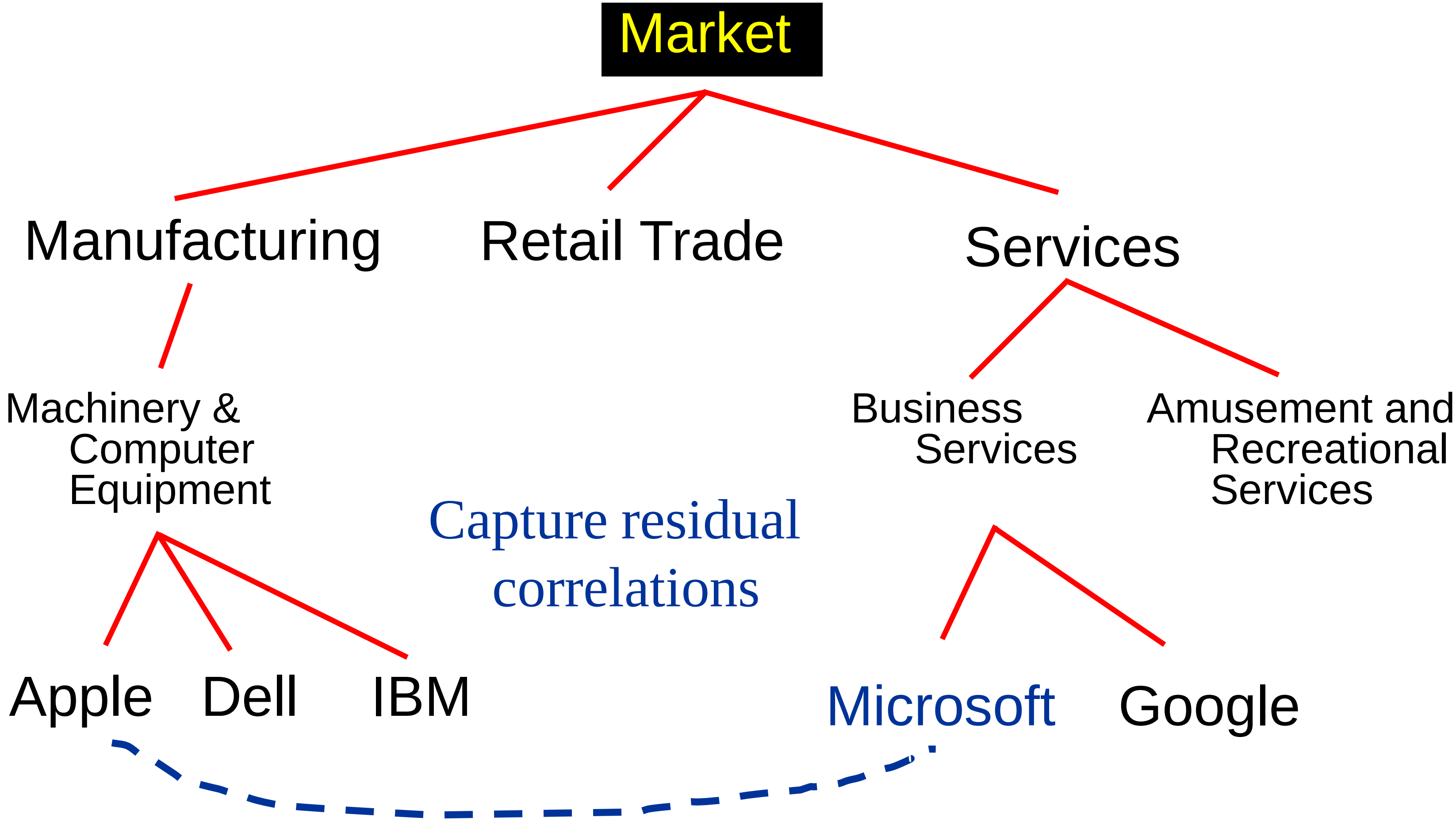
Multiresolution (MR) Tree Models



Multiresolution (MR) Tree Models



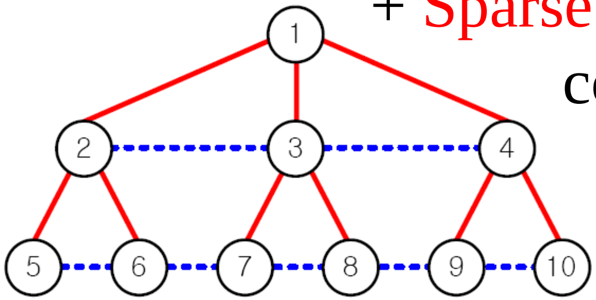
Multiresolution (MR) Tree Models



Sparse Markov and Covariance Structure

- MR tree model (sparse Markov structure)

+ Sparse covariance structure to capture residual correlations conditioned on other scales



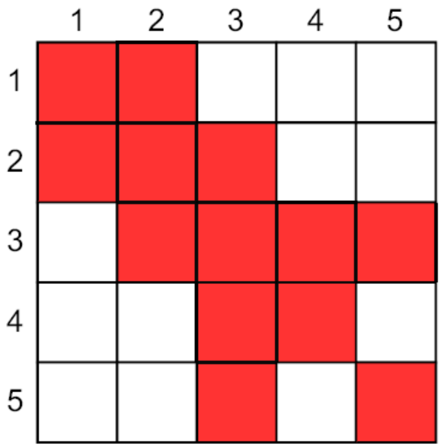
- Q1) Given noisy measurements at some of the nodes, how do we find the optimal estimate at all nodes?
- Q2) Given target covariance at the finest scale, how do we learn such a model?

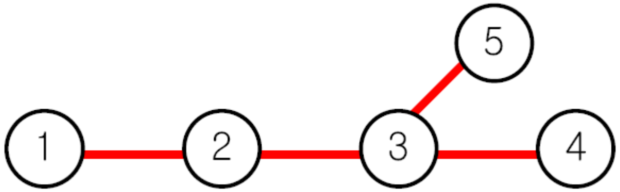
Gaussian Graphical Models

Gaussian Process $x \sim \mathcal{N}(\mu, \Sigma)$

Information Matrix $J = \Sigma^{-1}$

x is Markov with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E}): \quad J_{ij} \neq 0 \Leftrightarrow \{i, j\} \in \mathcal{E}$



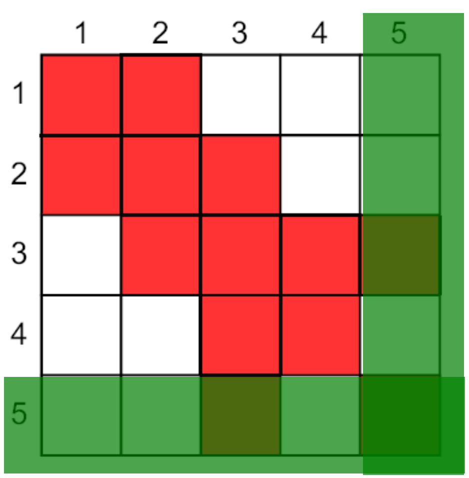


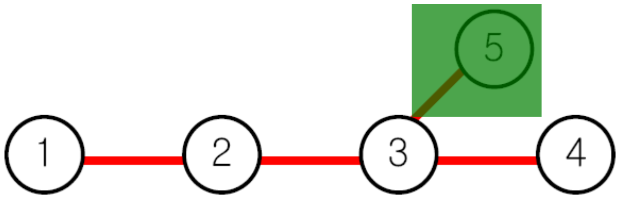
Gaussian Graphical Models

Gaussian Processes $x \sim \mathcal{N}(\mu, \Sigma)$

Information Matrix $J = \Sigma^{-1}$

x is Markov with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E}) : \; J_{ij} \neq 0 \Leftrightarrow \{i, j\} \in \mathcal{E}$

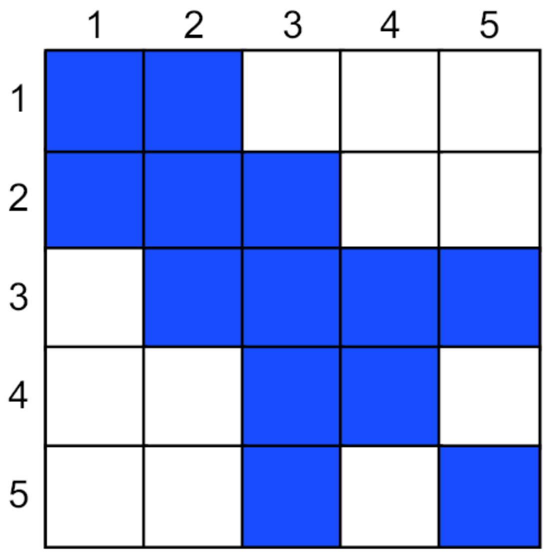




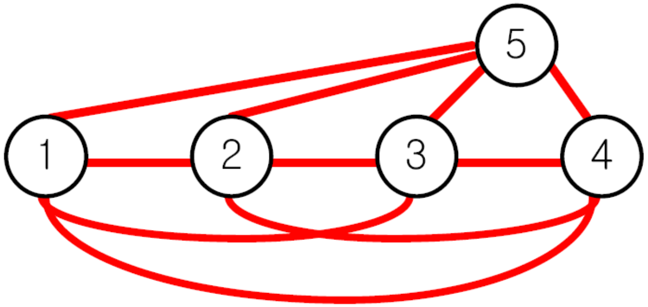
Conditional Distribution

p(x$_1$, x$_2$, x$_3$, x$_4$ | x$_5$)

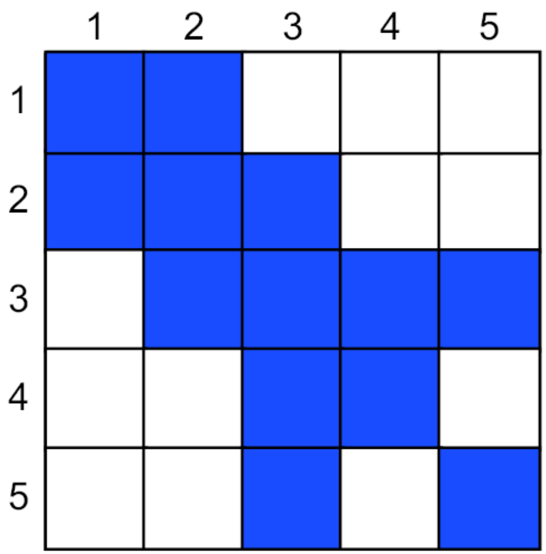
Conjugate Graphs



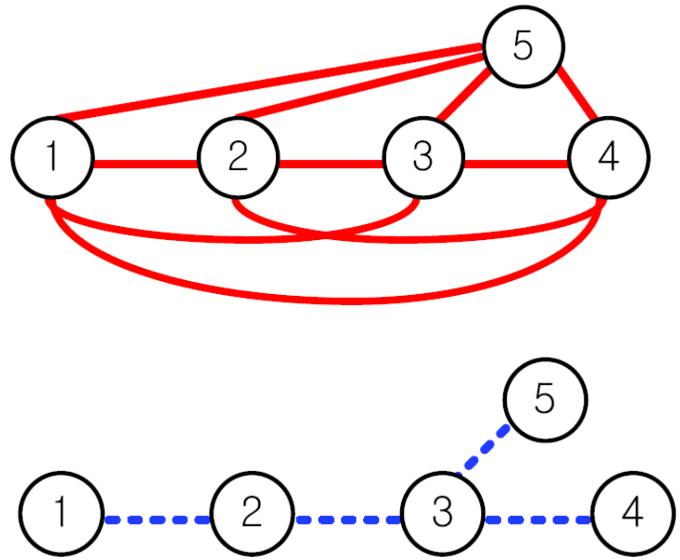
Sparsity of a
covariance matrix



Conjugate Graphs



Sparsity of a
covariance matrix



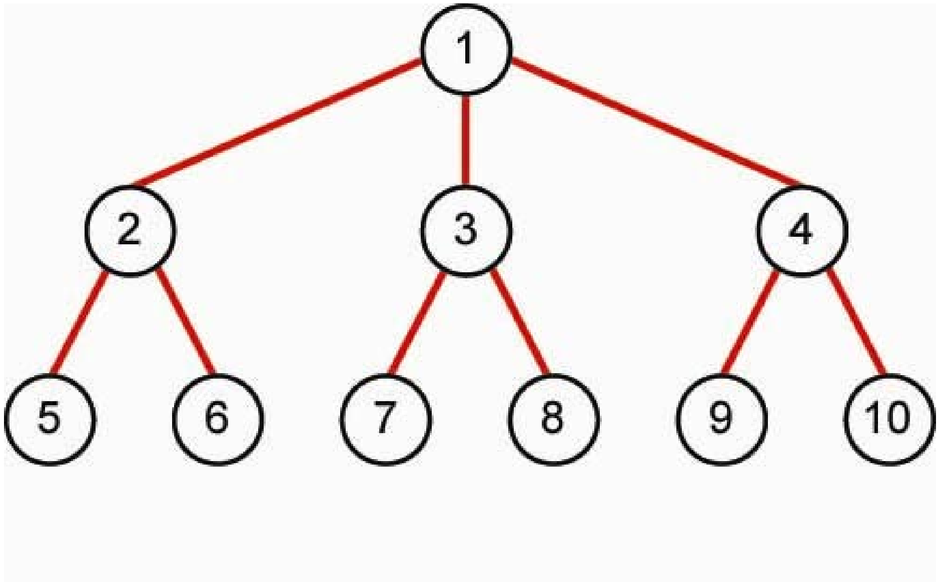
Conjugate edges in blue.

Sparse In-scale Conditional Covariance Multiresolution Model (SIM Model)

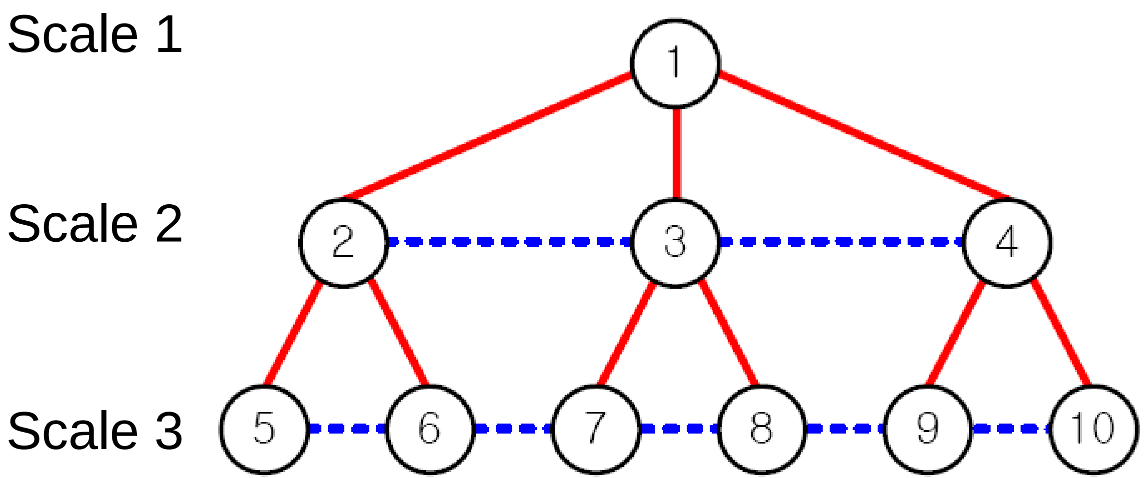
Scale 1

Scale 2

Scale 3

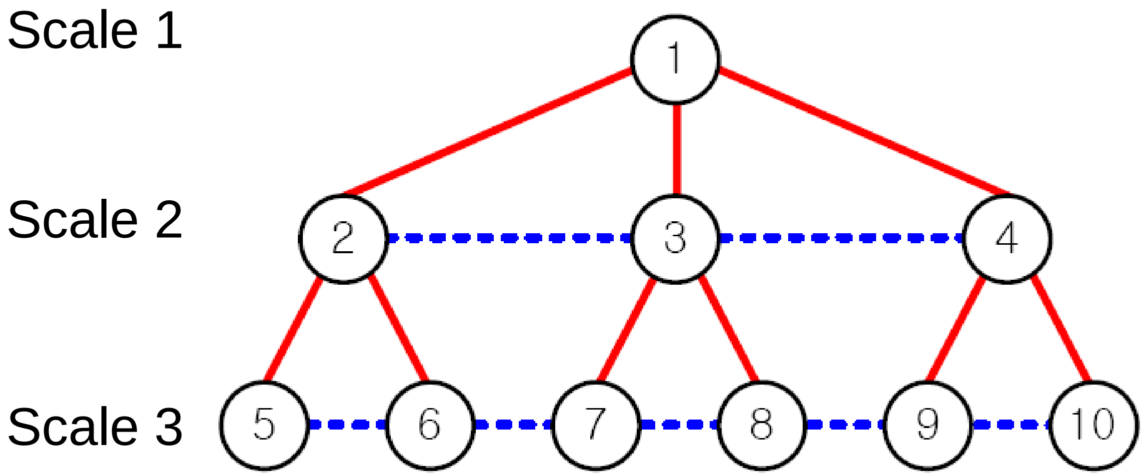


Sparse In-scale Conditional Covariance Multiresolution Model (SIM Model)

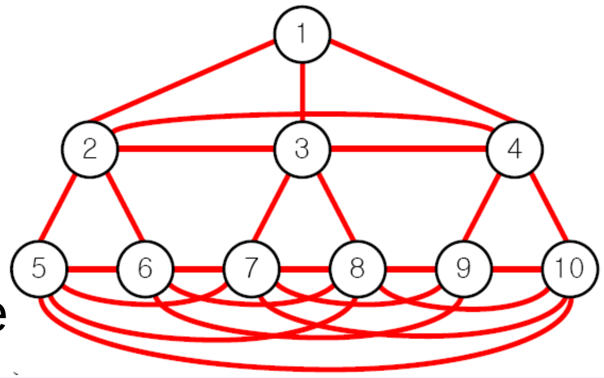


- Conditioned on scale 1 and scale 3, x_2 is independent to x_4.

Sparse In-scale Conditional Covariance Multiresolution Model (SIM Model)



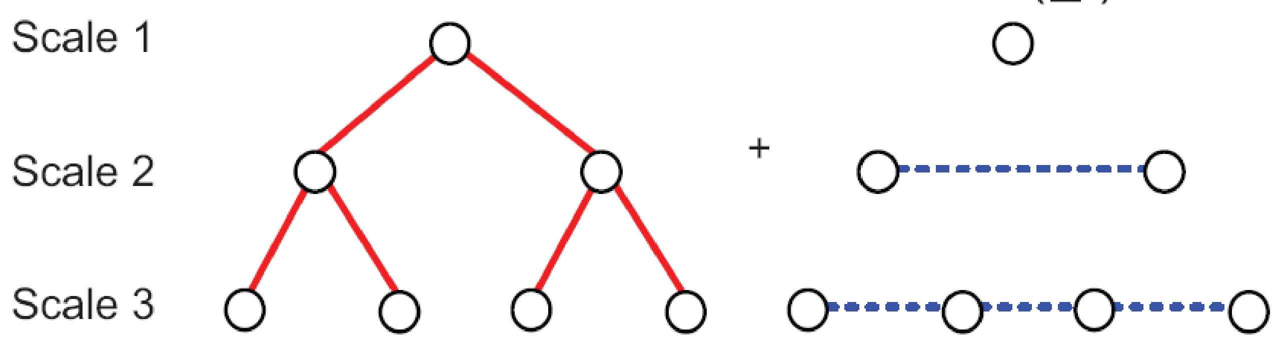
- Conditioned on scale 1 and scale 3, x_2 is independent to x_4.



Corresponding graphical model structure

Information Matrix of a SIM Model

$$J = \begin{pmatrix} J_{[1]} & J_{[1,2]} & 0 \\ J_{[2,1]} & J_{[2]} & J_{[2,3]} \\ 0 & J_{[3,2]} & J_{[3]} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & J_{[1,2]} & 0 \\ J_{[2,1]} & 0 & J_{[2,3]} \\ 0 & J_{[3,2]} & 0 \end{pmatrix}}_{J^h} + \underbrace{\begin{pmatrix} J_{[1]} & 0 & 0 \\ 0 & J_{[2]} & 0 \\ 0 & 0 & J_{[3]} \end{pmatrix}}_{J^c = (\Sigma^c)^{-1}}$$



Inference on Gaussian Models

Prior distribution of $x:\ \mathcal{N}(0, J^{-1})$

Noisy measurements: $y = Cx + v$

$$\hat{x} = \arg\max_x p(x|y) = (J + \underbrace{J^p}_{\equiv C^T R^{-1} C})^{-1} \underbrace{h}_{C^T R^{-1} y}$$

Solving $A\hat{x} = h$ iteratively by "matrix splitting"

$$A = M - K \Rightarrow \underbrace{M}_{\text{preconditioner}} \hat{x}^{new} = h + K\hat{x}^{old}$$

Inference in SIM Models

$$\boxed{(J^h + (\Sigma^c)^{-1} + J^p)\hat{x} = h}$$

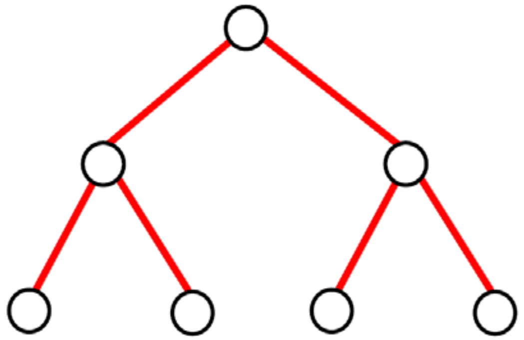
Inference in SIM Models

$$\boxed{\left(J^h + (\Sigma^c)^{-1} + J^p\right)\hat{x} = h}$$

$$J^p \equiv C^T R^{-1} C$$

- Tree Inference

$$\left(\underline{J^h + J^p \qquad}\right)\hat{x}^{new} = h - (\Sigma^c)^{-1}\hat{x}^{old}$$



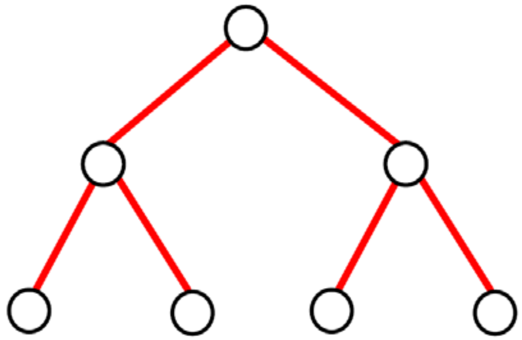
Inference in SIM Models

$$\boxed{(J^h + (\Sigma^c)^{-1} + \underline{J^p})\hat{x} = h}$$

$$\equiv C^T R^{-1} C$$

- Tree Inference

$$\underline{(J^h + J^p + D)}\hat{x}^{new} = h - (\Sigma^c)^{-1}\hat{x}^{old} + D\hat{x}^{old}$$



Computing $z \equiv (\Sigma^c)^{-1}\hat{x}$

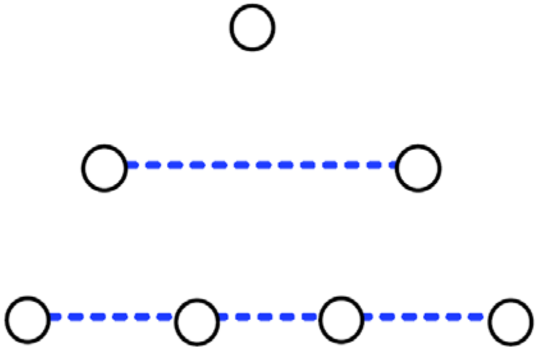
Solving $\underline{\Sigma^c} z = \hat{x}$

sparse, well-conditioned

Inference in SIM Models

- In-scale Inference

$$(\Sigma^c)^{-1}\,\hat{x}^{new} = \quad (h - J^h \hat{x}^{old} - J^p \hat{x}^{old})$$

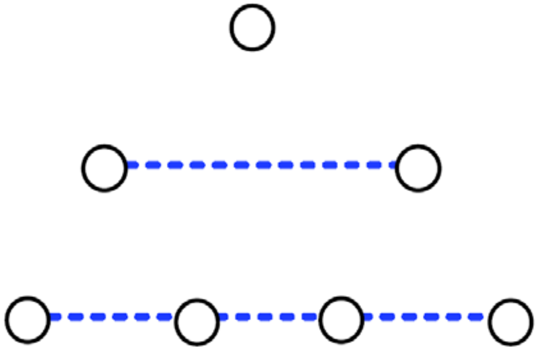


Inference in SIM Models

- In-scale Inference

$$\hat{x}^{new} = \underline{\Sigma}^{c}(h - \underline{J}^{h}\hat{x}^{old} - \underline{J}^{p}\hat{x}^{old})$$

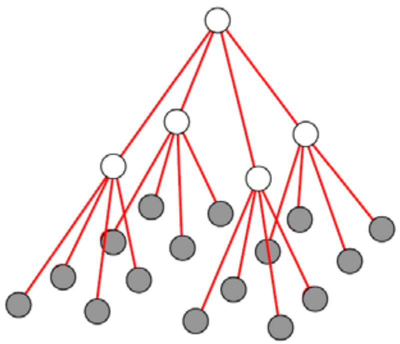
→ Sparse matrix multiplications



Learning a SIM model

Given the target covariance at the finest scale,

1. Learn an MR tree model.



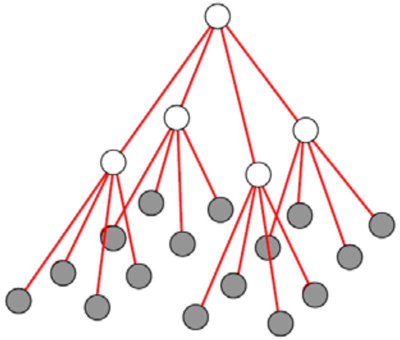
$$\begin{pmatrix} 0 & J_{[1,2]} & 0 \\ J_{[2,1]} & 0 & J_{[2,3]} \\ 0 & J_{[3,2]} & 0 \end{pmatrix} + \begin{pmatrix} \ddots & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \ddots \end{pmatrix}$$

2. Augment in-scale structures so that the marginal covariance at the finest scale exactly matches the target covariance.
3. Optimize the in-scale structure.

Learning a SIM model

Given the target covariance at the finest scale,

1. Learn an MR tree model.



$$\begin{pmatrix} 0 & J_{[1,2]} & 0 \\ J_{[2,1]} & 0 & J_{[2,3]} \\ 0 & J_{[3,2]} & 0 \end{pmatrix} + \begin{pmatrix} J_{[1]} & 0 & 0 \\ 0 & J_{[2]} & 0 \\ 0 & 0 & J_{[3]} \end{pmatrix}$$

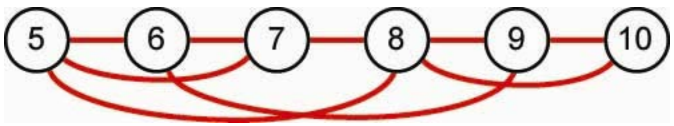
2. Augment in-scale structures so that the marginal covariance at the finest scale exactly matches the target covariance.
3. Optimize the in-scale structure.

Optimizing In-scale Structure

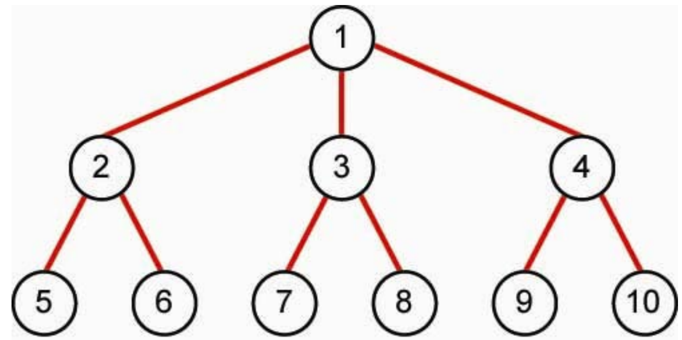
$$
\begin{aligned}
\hat{J} = \operatorname*{argmax}_{J \succ 0} \quad & \sum_m \log\det J_{[m]} \\
\text{s.t.} \quad & |J_{i,j} - J^*_{i,j}| \le \gamma_{i,j}, \;\; \forall \{i,j\} \in \mathcal{E}_{inscale} \\
& J_{i,j} - J^*_{i,j} = 0 \;\; \forall \{i,j\} \in \mathcal{E}_{inter}
\end{aligned}
$$

- Convex optimization problem
- $(\hat{J}_{[m]})^{-1}$ sparse.

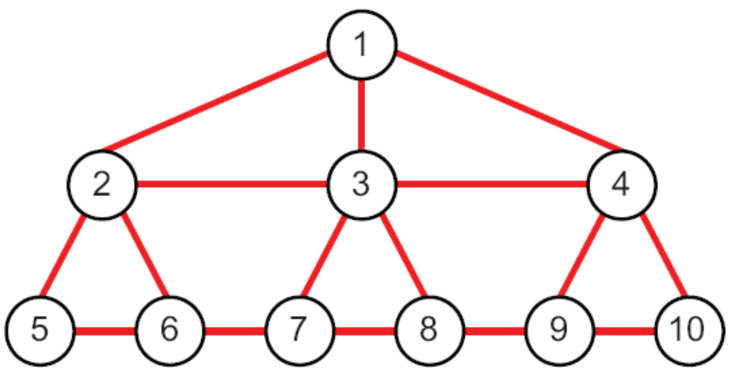
Experimental Results



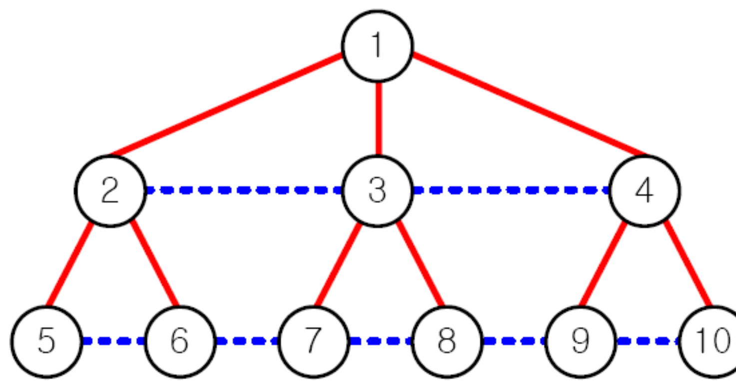
Single-scale model



MR tree model

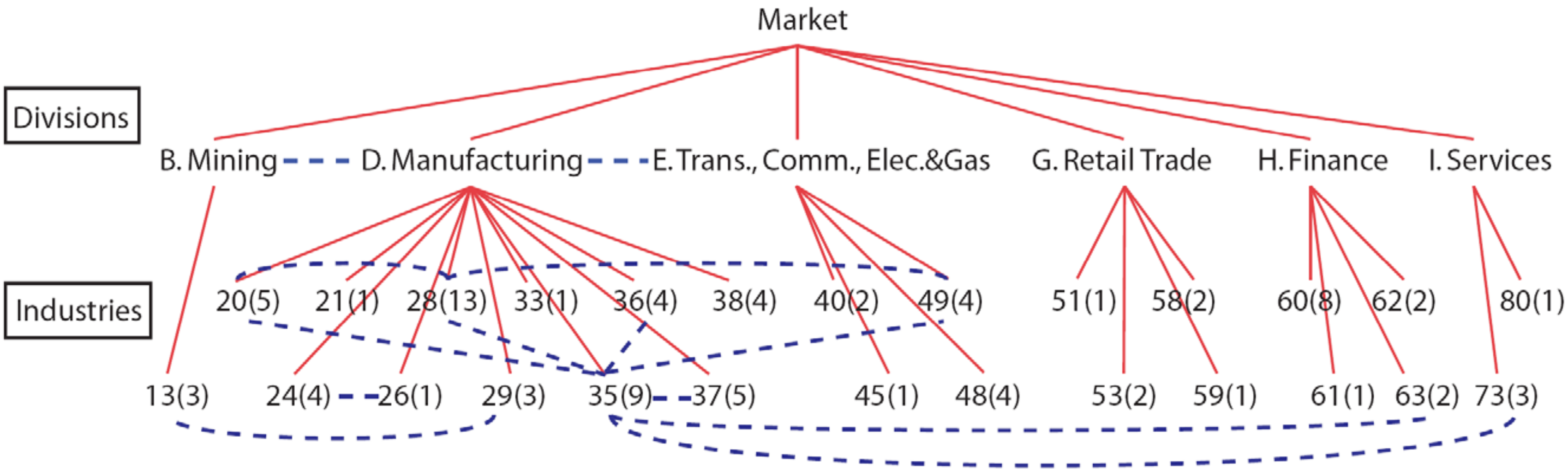


Sparse MR model



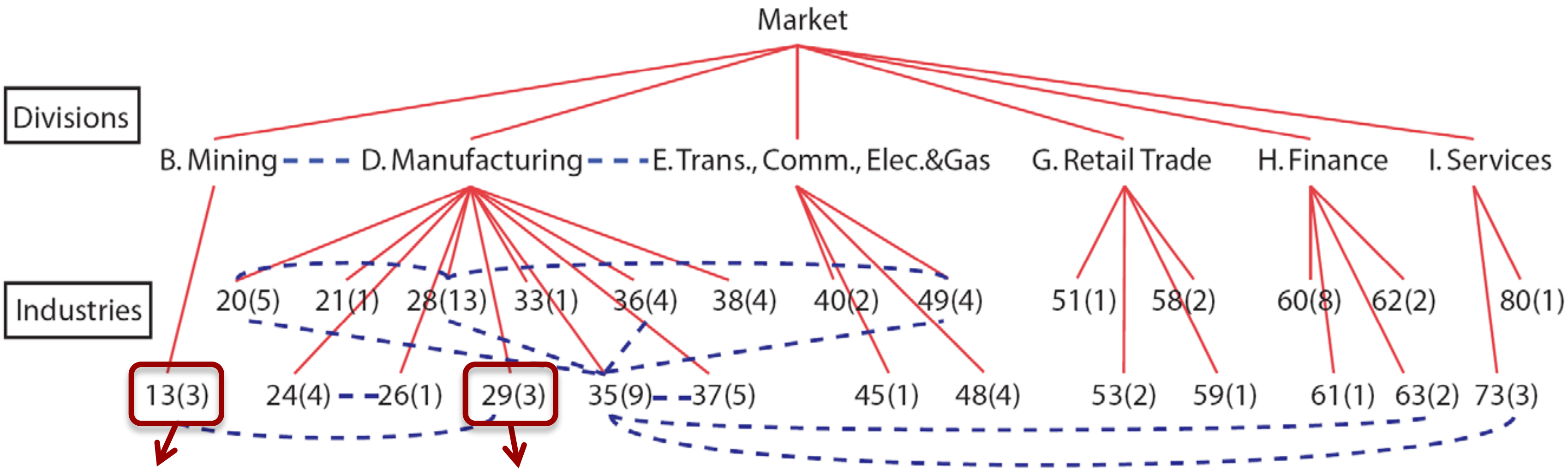
SIM model

Stock Returns Example



- Monthly returns of 84 companies in the S&P 100 index (1990-2007)
- Hierarchy based on the Standard Industrial Classification system
- Market, 6 divisions, 26 industries, and 84 individual companies

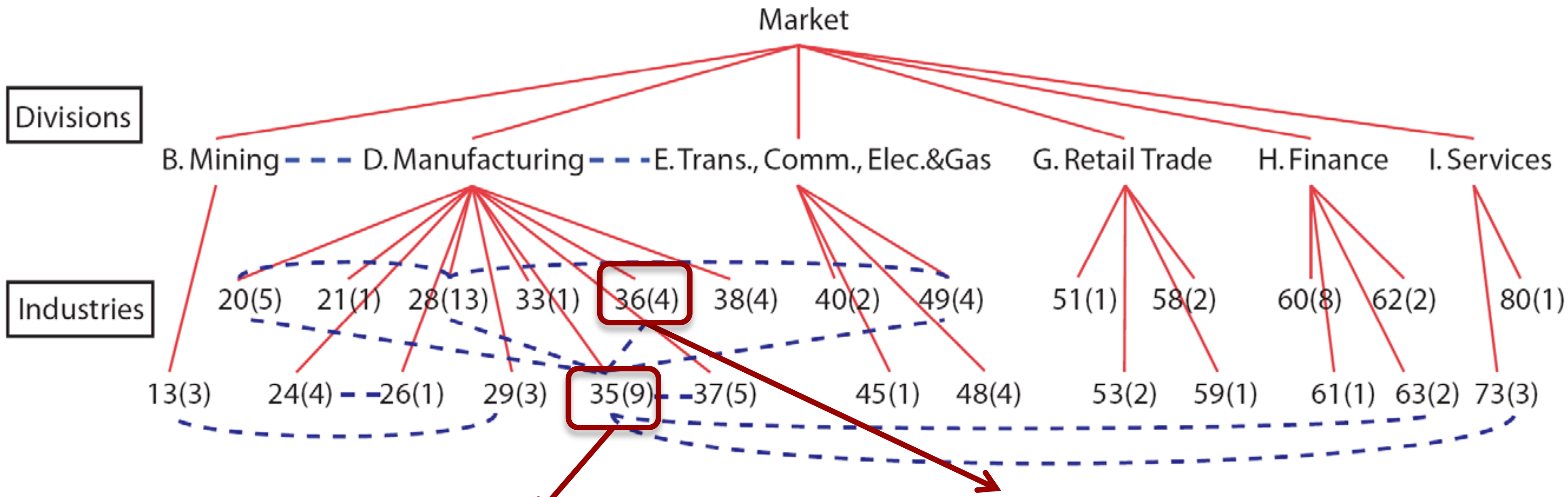
Stock Returns Example



Oil and Gas Extraction (Schlumberger)

Petroleum Refining (Exxon Mobile, Chevron)

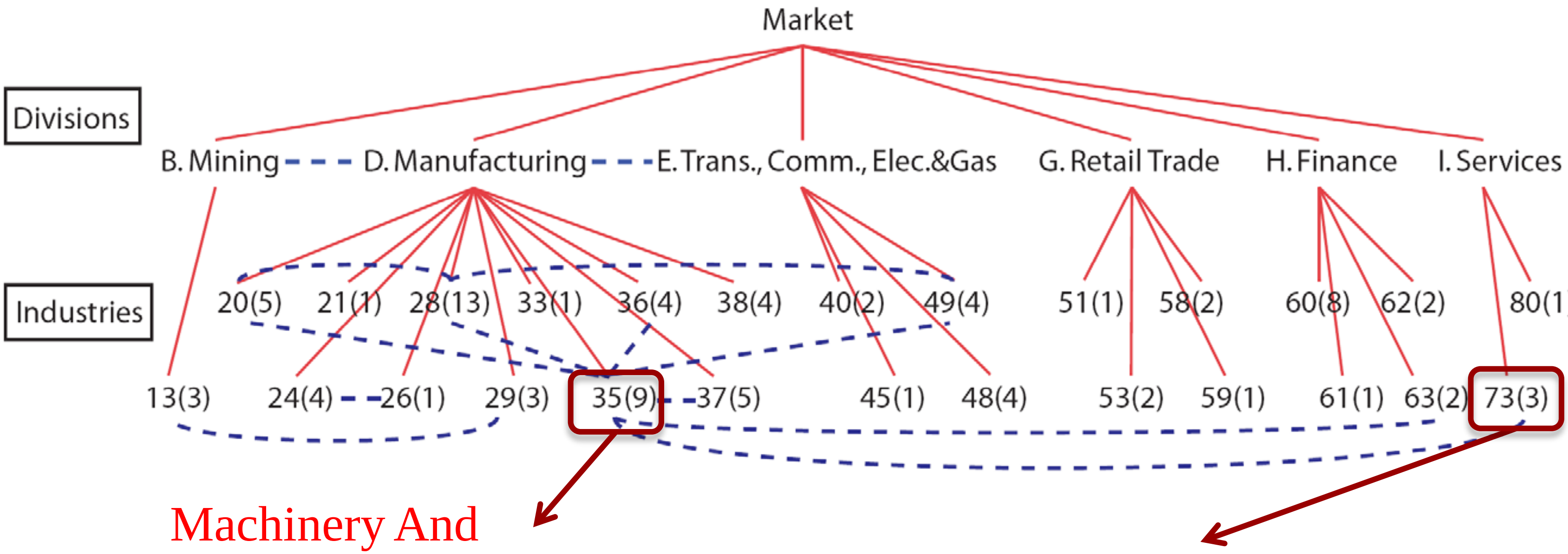
Stock Returns Example



Machinery And
Computer Equipment
(Dell, Apple, IBM, Xerox)

Other Electrical
Equipment
(TI, Intel, GE)

Stock Returns Example

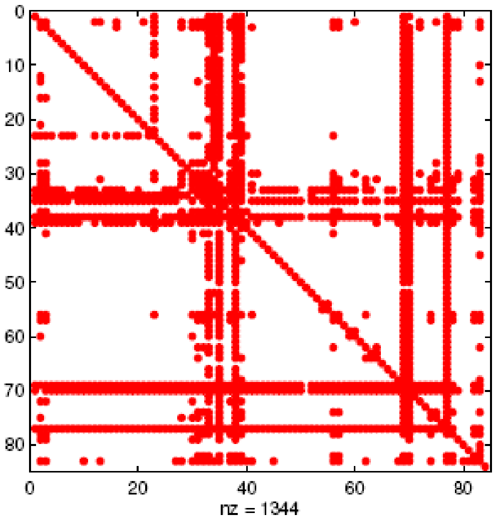


Machinery And Computer Equipment (Dell, Apple, IBM, Xerox)

Business Services (Microsoft, Oracle)

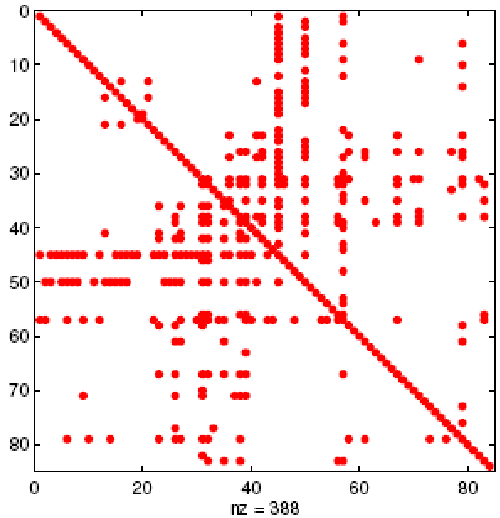
Stock Returns Example

- Sparsity Pattern -



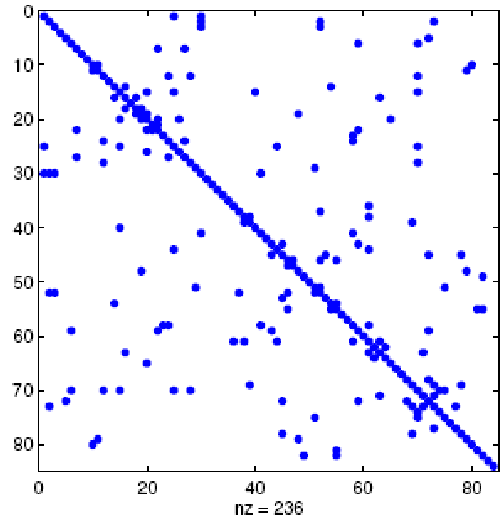
Single-scale

122.48



Sparse MR (finest scale)

28.34



SIM (finest scale)

16.36

Divergence between the approximate and the empirical distribution

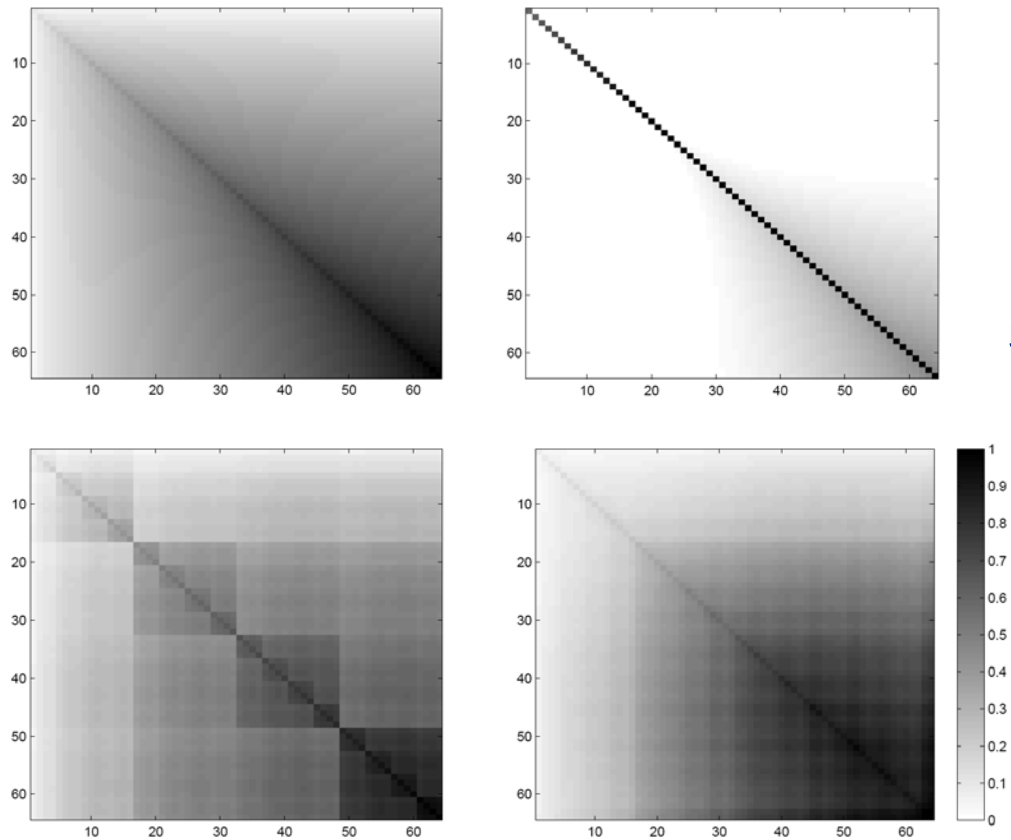
* Tree: 38.22

Fractional Brownian Motion

- Covariance Approximation -

Original

Single-scale

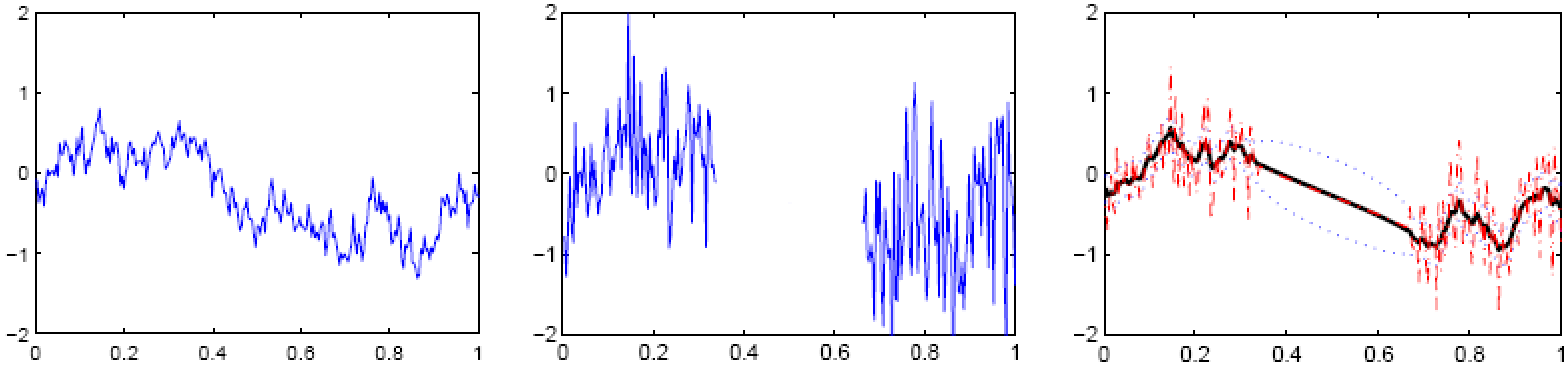


Tree

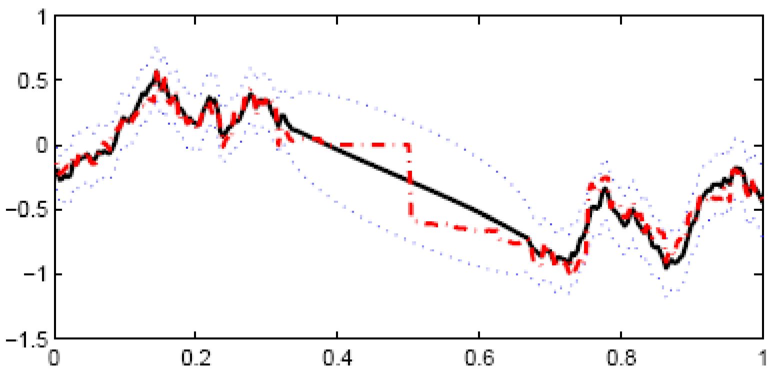
SIM

Fractional Brownian Motion

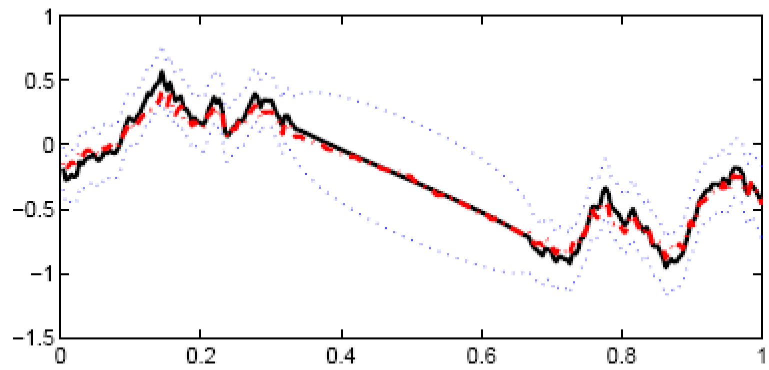
- Estimation -



Single-scale



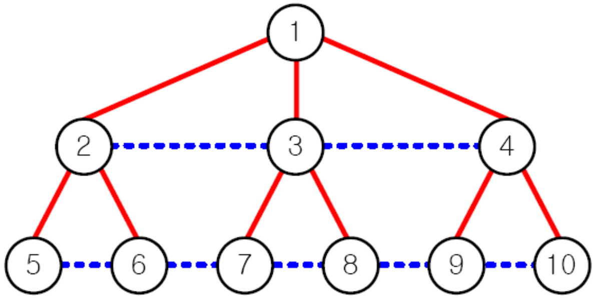
Tree



SIM

Conclusion and Future Work

- Sparse In-scale Conditional Covariance MR Models
  - Compact structure
  - Modeling and inference advantages

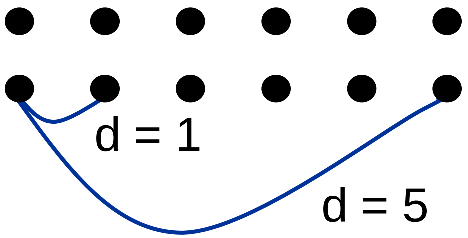


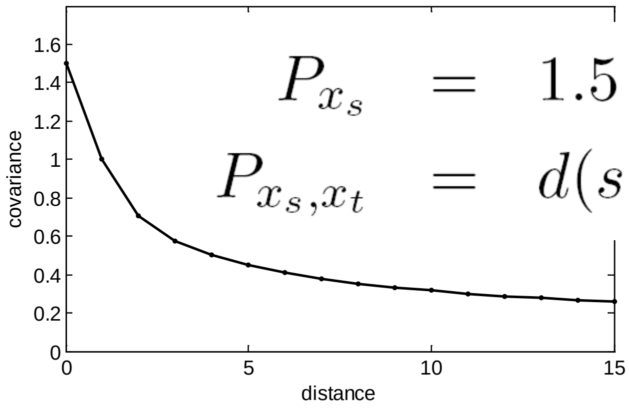
- Future work: extension to discrete models

Polynomially Decaying Covariance

- 256 Gaussian variables arranged spatially on a 16x16 grid.

- Covariance with polynomial decay:



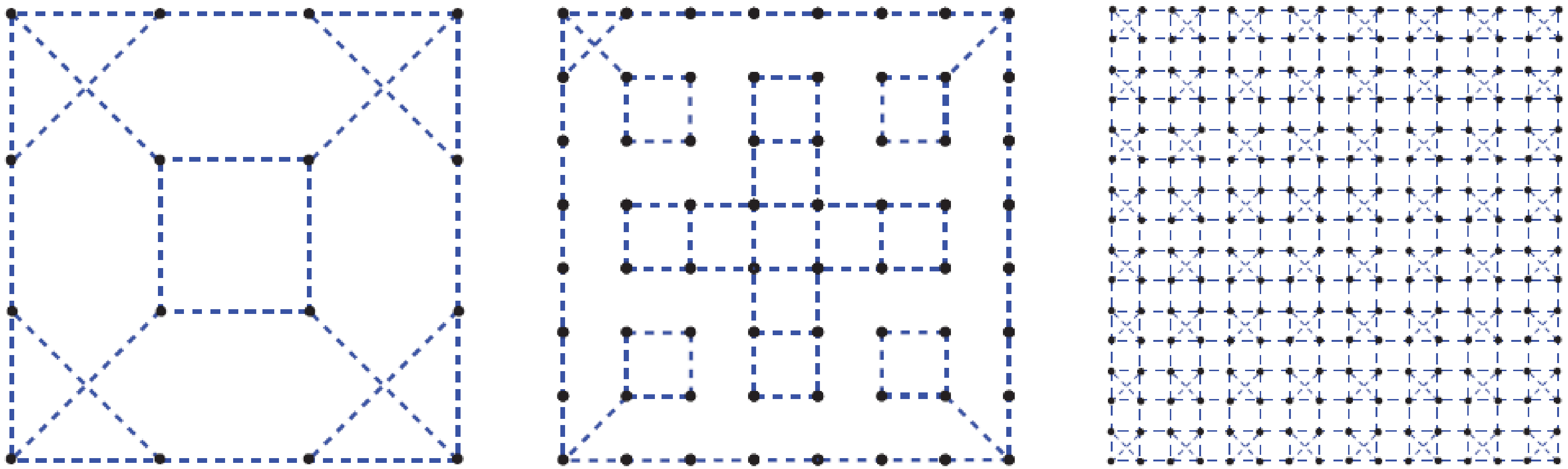


$$P_{x_s} = 1.5$$

$$P_{x_s, x_t} = d(s,t)^{-0.5}$$

Polynomially Decaying Covariance

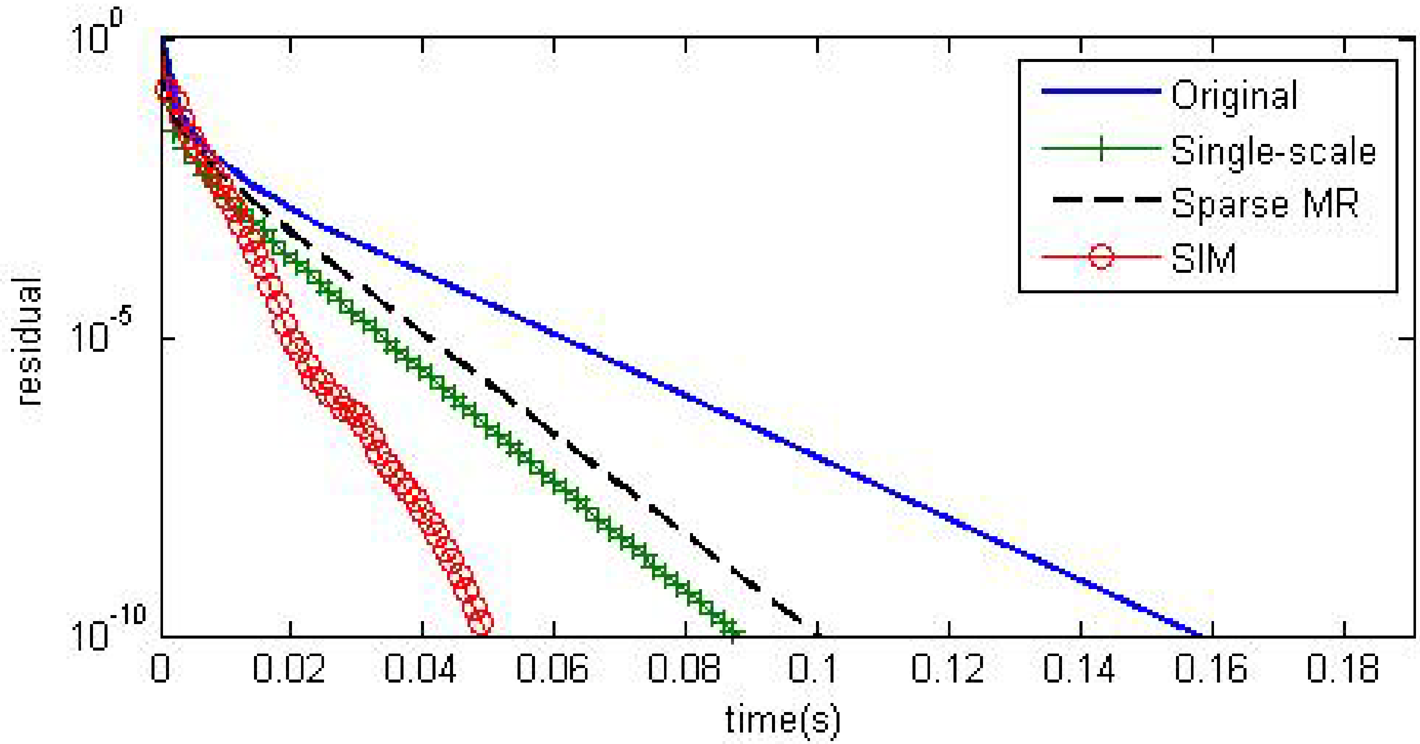
- Conjugate Graphs -



- Conjugate graph at each scale of the SCM model.
- Single-scale approximation densely connected (minimum degree 31).

Polynomially Decaying Covariance

- Estimation Performance -



- Estimation given sparse noisy measurements.
- Residual error vs. computation time.