Growing money on Trees & Graphs

Stock market analysis using Multidimensional Scaling

Ravee Malla

2008CS50224
CSC 410 - Colloquium
IIT Delhi

October 11 2011

Motivation

- Economics is too serious a subject to be left to *economists!*
- Identify trends and clusters in the market
- Study time evolution of the market
- Investor risk reduction
- Predict future trends
- Ease of visualization

Economics preliminaries

- **Stock:** A notional unit of the ownership of a company

Economics preliminaries

- **Stock:** A notional unit of the ownership of a company
- **Market:** A set of listed companies, denoted by $\mathcal{C}$

Economics preliminaries

- **Stock:** A notional unit of the ownership of a company
- **Market:** A set of listed companies, denoted by $\mathcal{C}$
- **Stock price:** $p : \mathcal{C} \times T \to \mathbb{R}_+$

Economics preliminaries

- **Stock:** A notional unit of the ownership of a company
- **Market:** A set of listed companies, denoted by $\mathcal{C}$
- **Stock price:** $p : \mathcal{C} \times T \to \mathbb{R}_+$
- **Return on stock:** $r : p \times \mathcal{C} \times T \to \mathbb{R} := \ln \frac{p(c,\tau)}{p(c,\tau-1)}$

Economics preliminaries

- **Stock:** A notional unit of the ownership of a company
- **Market:** A set of listed companies, denoted by $\mathcal{C}$
- **Stock price:** $p : \mathcal{C} \times T \to \mathbb{R}_+$
- **Return on stock:** $r : p \times \mathcal{C} \times T \to \mathbb{R} := \ln \frac{p(c,\tau)}{p(c,\tau-1)}$
- **Portfolio:** A set $\mathcal{P} = \langle \mathcal{B}, p, r, \mathcal{N}, \tau \rangle$ where $\mathcal{B} \subset \mathcal{C}$, $\mathcal{N} : \mathcal{B} \to \mathbb{Z}_+$

Problem Overview

- **Portfolio Optimization:** Design a portfolio $\mathcal{P}$ s.t.

$$\begin{aligned} \underset{\mathcal{B} \subset \mathcal{C}}{\text{maximize}} \quad & \sum_{b \in \mathcal{B}} r(p, b, \tau) * N(b) \\ \text{subject to} \quad & \sum_{b \in \mathcal{B}} p(b, \tau) * N(b) \leq \Gamma \end{aligned} \tag{1}$$

Problem Overview

- **Portfolio Optimization:** Design a portfolio $\mathcal{P}$ s.t.

$$\begin{aligned} \underset{\mathcal{B} \subset \mathcal{C}}{\text{maximize}} \quad & \sum_{b \in \mathcal{B}} r(p, b, \tau) * N(b) \\ \text{subject to} \quad & \sum_{b \in \mathcal{B}} p(b, \tau) * N(b) \leq \Gamma \end{aligned} \tag{1}$$

- **Diversification:** How to invest in independent sectors to reduce overall risk

Problem Overview

- **Portfolio Optimization:** Design a portfolio $\mathcal{P}$ s.t.

$$\begin{aligned} \underset{\mathcal{B} \subset \mathcal{C}}{\text{maximize}} \quad & \sum_{b \in \mathcal{B}} r(p, b, \tau) * N(b) \\ \text{subject to} \quad & \sum_{b \in \mathcal{B}} p(b, \tau) * N(b) \leq \Gamma \end{aligned} \tag{1}$$

- **Diversification:** How to invest in independent sectors to reduce overall risk
- **Create a tool to help analyze the market and make better investment decisions**

What is Multidimensional Scaling?

- Statistical technique to explore similarities in data and visualize on a map

What is Multidimensional Scaling?

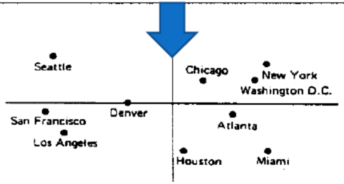
- Statistical technique to explore similarities in data and visualize on a map
- Given distances d_{ij} on $\mathcal{C}^2$, learn a function $f : \mathcal{C} \to \mathbb{R}^2$

What is Multidimensional Scaling?

- Statistical technique to explore similarities in data and visualize on a map
- Given distances d_{ij} on $\mathcal{C}^2$, learn a function $f : \mathcal{C} \rightarrow \mathbb{R}^2$

Table 1 Flying Mileages Between 10 American Cities

Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington, DC	
0	587	1212	701	1936	604	748	2139	2182	543	Atlanta
587	0	920	940	1745	1188	713	1858	1737	597	Chicago
1212	920	0	879	831	1726	1631	949	1021	1494	Denver
701	940	879	0	1374	968	1420	1645	1891	1220	Houston
1936	1745	831	1374	0	2339	2451	347	959	2300	Los Angeles
604	1188	1726	968	2339	0	1092	2594	2734	923	Miami
748	713	1631	1420	2451	1092	0	2571	2408	205	New York
2139	1858	949	1645	347	2594	2571	0	678	2442	San Francisco
2182	1737	1021	1891	959	2734	2408	678	0	2329	Seattle
543	597	1494	1220	2300	923	205	2442	2329	0	Washington, DC



A map of the 10 cities

Measure on the market

- MDS requires a metric on the entity set

Measure on the market

- MDS requires a metric on the entity set
- Given the *state* of the market, define a measure
  - Model complex interactions and temporal behaviour
  - Help in finding clusters and trends in the market
  - Well behaved, easy to compute and bounded

Measure on the market

- MDS requires a metric on the entity set
- Given the *state* of the market, define a measure
    - Model complex interactions and temporal behaviour
    - Help in finding clusters and trends in the market
    - Well behaved, easy to compute and bounded
- What can this measure be?

Measure on the market

- MDS requires a metric on the entity set
- Given the *state* of the market, define a measure
  - Model complex interactions and temporal behaviour
  - Help in finding clusters and trends in the market
  - Well behaved, easy to compute and bounded
- What can this measure be? **Pearson correlation**
- $\rho : \mathcal{C} \times \mathcal{C} \times T \rightarrow [-1, 1] \equiv \rho^{\tau}_{i,j} \rightarrow \frac{\langle r_i^{\tau} r_j^{\tau} \rangle - \langle r_i^{\tau} \rangle \langle r_j^{\tau} \rangle}{\sqrt{[\langle {r_i^{\tau}}^2 \rangle - \langle r_i^{\tau} \rangle^2][\langle {r_j^{\tau}}^2 \rangle - \langle r_j^{\tau} \rangle^2]}}$

Measure on the market

- MDS requires a metric on the entity set
- Given the *state* of the market, define a measure
  - Model complex interactions and temporal behaviour
  - Help in finding clusters and trends in the market
  - Well behaved, easy to compute and bounded
- What can this measure be? **Pearson correlation**
- $\rho : \mathcal{C} \times \mathcal{C} \times T \to [-1, 1] \equiv \rho^{\tau}_{i,j} \to \frac{\langle r_i^{\tau} r_j^{\tau} \rangle - \langle r_i^{\tau} \rangle \langle r_j^{\tau} \rangle}{\sqrt{[\langle r_i^{\tau\,2} \rangle - \langle r_i^{\tau} \rangle^2][\langle r_j^{\tau\,2} \rangle - \langle r_j^{\tau} \rangle^2]}}$
- Why?
  - Globally high ρ implies that market is generally correlated, so it is either in a recession or boom phase
  - Treasure trove of rich probability theory applicable
  - Easy to compute and analyze

Metric on the market

- Given the correlation measure ρ, define $d_{ij} = \sqrt{2(1 - \rho_{ij})}$
- Properties preserved by $d, \forall i, j, k$
  - $d_{ij} \geq 0$
  - $d_{ii} = 0$
  - $d_{ij} < d_{ik} + d_{kj}$
  - $d_{ij} = d_{ji}$
  - Ultrametricity: Mapping preserves linearity

Using MDS

- Consider company distances d^{τ} as our distance metric
- Randomly assign a vector $X_c^{\tau} = (x_c^{\tau}, y_c^{\tau})$ to each company c

Using MDS

- Consider company distances d^τ as our distance metric
- Randomly assign a vector $X_c^\tau = (x_c^\tau, y_c^\tau)$ to each company c
- Define stress $S^\tau(X_1^\tau, ..., X_N^\tau) = \sum_{i \neq j} \sqrt{{d_{i,j}^\tau}^2 - \|X_i^\tau - X_j^\tau\|^2}$
- There on, it's just an optimization problem!

Using MDS

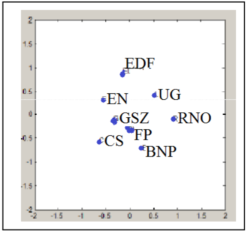
- Consider company distances d^τ as our distance metric
- Randomly assign a vector $X_c^\tau = (x_c^\tau, y_c^\tau)$ to each company c
- Define stress $S^\tau(X_1^\tau, \ldots, X_N^\tau) = \sum_{i \neq j} \sqrt{{d_{i,j}^\tau}^2 - \|X_i^\tau - X_j^\tau\|^2}$
- There on, it's just an optimization problem!
- Why just 2 dimensions?

Using MDS

- Consider company distances d^τ as our distance metric
- Randomly assign a vector $X_c^\tau = (x_c^\tau, y_c^\tau)$ to each company c
- Define stress $S^\tau(X_1^\tau, ..., X_N^\tau) = \sum_{i \neq j} \sqrt{{d_{i,j}^\tau}^2 - \|X_i^\tau - X_j^\tau\|^2}$
- There on, it's just an optimization problem!
- Why just 2 dimensions?
  - Visualization of the market
  - Stress doesn't decrease significantly in higher dimensions

Using MDS

- Consider company distances d^τ as our distance metric
- Randomly assign a vector $X_c^\tau = (x_c^\tau, y_c^\tau)$ to each company c
- Define stress $S^\tau(X_1^\tau, ..., X_N^\tau) = \sum_{i \neq j} \sqrt{{d_{i,j}^\tau}^2 - \|X_i^\tau - X_j^\tau\|^2}$
- There on, it's just an optimization problem!
- Why just 2 dimensions?
  - Visualization of the market
  - Stress doesn't decrease significantly in higher dimensions



Time evolution more interesting!

- Generate time evolving maps to reflect how market changes

Time evolution more interesting!

- Generate time evolving maps to reflect how market changes
- Easy, apply MDS algorithm to d^{τ} for successive τ

Time evolution more interesting!

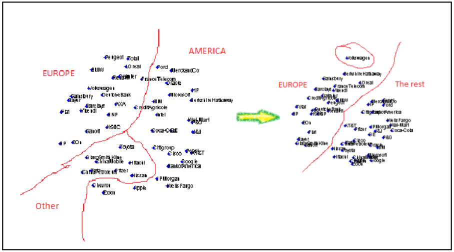
- Generate time evolving maps to reflect how market changes
- Easy, apply MDS algorithm to d^τ for successive τ
- **Critical Flaw!** MDS algorithm assumes random initialization, so successive maps will change dramatically.

Time evolution more interesting!

- Generate time evolving maps to reflect how market changes
- Easy, apply MDS algorithm to d^τ for successive τ
- **Critical Flaw!** MDS algorithm assumes random initialization, so successive maps will change dramatically.
- Add another component to S penalizing changes in assigned coordinates
- $S^\tau(X_1^\tau, \ldots, X_N^\tau) = \sum_{i \neq j} \sqrt{{d_{i,j}^\tau}^2 - \|X_i^\tau - X_j^\tau\|^2} + \sum_i w_i * \|X_i^\tau - X_i^{\tau-1}\|$

Subprime crisis 2008

- 3 individual markets got correlated & came together due to a recession
- Can define global market score looking at the maps and learn it to predict future recession



Asset Tree

- Let $G^{\tau} = (\mathcal{C}, \mathcal{C} \times \mathcal{C}, w)$ where $w(i,j) = d^{\tau}(i,j)$

Asset Tree

- Let $G^\tau = (\mathcal{C}, \mathcal{C} \times \mathcal{C}, w)$ where $w(i,j) = d^\tau(i,j)$
- Consider a minimum spanning tree T^τ of G^τ in $\mathcal{O}(n+m)$

Asset Tree

- Let $G^\tau = (\mathcal{C}, \mathcal{C} \times \mathcal{C}, w)$ where $w(i,j) = d^\tau(i,j)$
- Consider a minimum spanning tree T^τ of G^τ in $\mathcal{O}(n+m)$
- Define a special node v_c on the tree based on *importance*
  - *highest vertex degree*
  - *most strongly correlated*

Asset Tree

- Let $G^\tau = (\mathcal{C}, \mathcal{C} \times \mathcal{C}, w)$ where $w(i,j) = d^\tau(i,j)$
- Consider a minimum spanning tree T^τ of G^τ in $\mathcal{O}(n+m)$
- Define a special node v_c on the tree based on *importance*
  - *highest vertex degree*
  - *most strongly correlated*
- Mean occupation layer $l(\tau, v_c) = \frac{1}{N} \sum_i depth(v_i^\tau)$

Asset Tree

- Let $G^\tau = (\mathcal{C}, \mathcal{C} \times \mathcal{C}, w)$ where $w(i,j) = d^\tau(i,j)$
- Consider a minimum spanning tree T^τ of G^τ in $\mathcal{O}(n+m)$
- Define a special node v_c on the tree based on *importance*
  - *highest vertex degree*
  - *most strongly correlated*
- Mean occupation layer $l(\tau, v_c) = \frac{1}{N} \sum_i depth(v_i^\tau)$
- Center of mass $v_m = \{v | v \in V, l(\tau, v) \text{ is min}\}$

Bibliography

- J. Tenreiro Machado, Fernando B. Duarte, Gonçalo Monteiro Duarte
  Multidimensional Scaling Analysis of Stock Market Values
- Ingwer Borg, Patrick J. F. Groenen
  Modern multidimensional scaling: theory and applications
- Warren S. Torgerson
  Multidimensional scaling: I. Theory and method
- A. Chakraborti, I.M Toke, M. Patriarca, F. Abergel
  Econophysics: Empirical facts and agent-based models
- J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertsz, A. Kanto
  Dynamics of market correlations: Taxonomy and portfolio analysis

Conclusions

Thanks

Questions or Comments?