Stock Portfolios - Taxonomy & Theory

Chapter 12: Correlation coefficient ρ_{ij} stocks i, j

ρ_{ij} useful because:

(i) allows us to define a metric for relative "distance" between stocks

(ii) provides mechanism for extracting economic information stored in stock-price time series

Let $S_i = \ln(Y_i(t)) - \ln(Y_i(t-1))$

$\rightarrow$ log of stock price difference

Define $$\tilde{S}_i \equiv \frac{S_i - \langle S_i \rangle}{\sqrt{\langle S_i^2 \rangle - \langle S_i \rangle^2}}$$

(a normalized log price difference. $\tilde{S}_i$ is also centered - i.e., mean = 0)

Let there be n elements in each stock time series (i.e., n times)

Then we can regard each $\tilde{S}_i$ as an n-element vector.

The Euclidean distance between vectors $\tilde{S}_i$ and $\tilde{S}_j$ is then:

$$d_{ij}^2 = \| \tilde{S}_i - \tilde{S}_j \| = \sum_{k=1}^{n} (\tilde{S}_{ik} - \tilde{S}_{jk})^2$$

Note that $\tilde{S}_i$ is in fact a unit vector because it was constructed so that

$$\sum_{k=1}^{n} \tilde{S}_{ik}^2 = 1$$ (Proof left as exercise)

So $$d_{ij}^2 = \sum_{k=1}^{n} (\tilde{S}_{jk}^2 + \tilde{S}_{ik}^2 - 2\tilde{S}_{ik}\tilde{S}_{jk})$$

$$= 2 - 2 \underbrace{\sum_{k=1}^{n} \tilde{S}_{ik}\tilde{S}_{jk}}$$

Note: $= \rho_{ij}$, correlation coefficient

(*) Thus $d_{ij} = \sqrt{2(1-\rho_{ij})}$

Since d_{ij} defines a Euclidean distance, following properties hold:

(i) $d_{ij} = 0$ if $i = j$ (because $\rho_{ii} = 0$)

(ii) $d_{ij} = d_{ji}$ (because $\rho_{ij} = \rho_{ji}$)

(iii) $d_{ij} \leq \underbrace{d_{ik} + d_{kj}}_{\text{components of } d_{ij}}$ standard triangular inequality

k index runs from $1 \Rightarrow n$, where k is a time index

Thus d_{ij} satisfies all properties of a metric space. This definition was due to Mantegna (1999).

We want to decompose a set of N objects into subsets of closely related objects

Ultrametric Spaces

A space defined by an ultrametric distance. Such a distance $\hat{d}_{ij}$ must satisfy (i) & (ii). However, (iii) is replaced by the stronger inequality, called an ultrametric inequality:

$$\hat{d}_{ij} \leq \max\{\hat{d}_{ik}, \hat{d}_{kj}\}$$

$\hat{d}_{ij}$ can be used to define a heirarchical system such as a spin glass. See Rammal et al (1986).

Idea is to use metric distance $\hat{d}_{ij}$ to partition the N objects.

Subdominant ultrametric: Obtained by determining The Minimal Spanning Tree connecting the N objects.

Graph Theory: In a connected weighted graph of N objects, the MST is a tree graph having

N-1 edges that minimizes the sum of the edge distances.

⟹ Subdominant ultrametric space provides a unique indexed heirarchy

⟹ Kruskal's algorithm is often used to construct the MST.

Kruskal is a sorting algorithm based on d_{ij}

$\boxed{\text{See Chapter 13 of Mantegna \& Stanley for further details}}$

Note: This chapter shows how to group stocks based on similarities

- But good portfolios select stocks that have significant differences

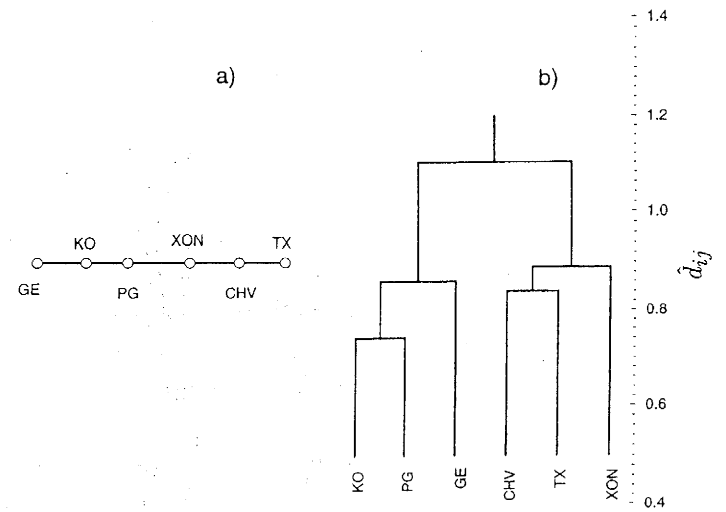

Fig. 13.1. (a) MST and (b) indexed hierarchical tree obtained for the example of firms, identified by their tick symbols CHV, GE, KO, PG, TX and XON.

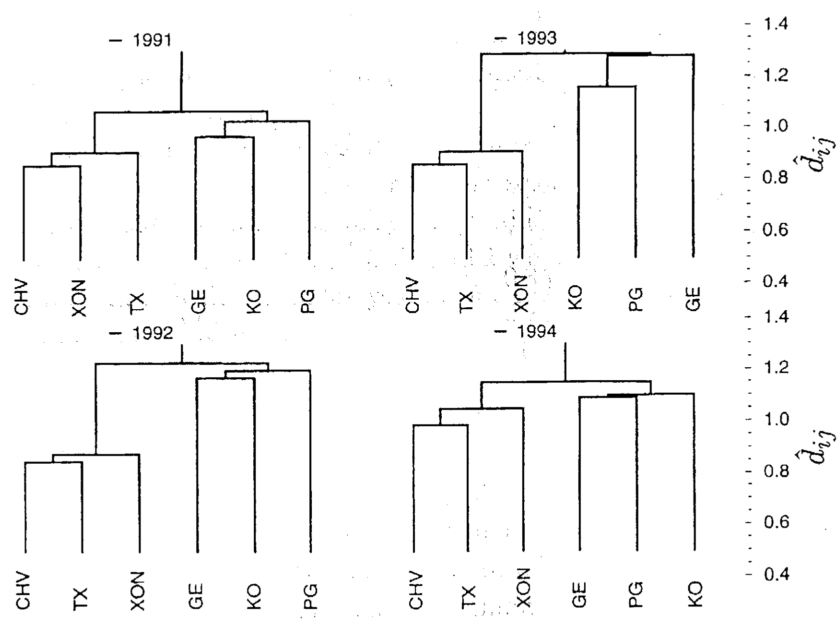

Fig. 13.2. Indexed hierarchical trees obtained during the calendar years from 1991 to 1994 for the portfolio of six firms (CHV, GE, KO, PG, TX, and XON).

	CHV	GE	KO	PG	TX	XON
CHV	0	1.10	1.10	1.10	0.84	0.89
GE		0	0.86	0.86	1.10	1.10
KO			0	0.74	1.10	1.10
PG				0	1.10	1.10
TX					0	0.89
XON						0