Hierarchical Structure in Financial Markets

Abstract Summary-

The paper aims to find a way to present stocks of a portfolio in a topological arrangement or in other words as vertices of a weighted graph with an appropriate metric for distance.
The graph is obtained from the matrix of correlation coefficient between daily logarithmic returns computed between all pairs of stocks of the portfolio.
This correlation coefficient is then used to calculate an appropriate metric to determine the “closeness” or weight of the edge between two vertices.
It then uses the minimum spanning tree of this graph to prove the existence and study the nature of common economic factors which affect the prices of these stocks.
Finally,the hierarchical tree of the subdominant ultrametric space associated with the graph obtained from the minimum spanning tree is studied to group stocks together which are affected by common economic factors.

Keywords and definitions-

1)Topological arrangement :

To express stocks on paper as a graph with weights of edges decided by an appropriate metric which satisfies all of the euclidean axioms.

2)Minimum Spanning tree :

A spanning tree is a subgraph wherein all the vertices are connected by the least possible number of edges. Of all possible spanning trees, the one in which the sum of weights of edges is minimum is the minimum spanning tree.

3) Hierarchical tree of the subdominant ultrametric space-

It is a graph with stocks on one axis and the distance between them on the other axis. The distance between 2 stocks is the maximum value of the metric distance detected by moving in single steps from one stock to another through the path connecting the two in the MST.

4)Correlation coefficient -

It measures how one quantity changes with respect to the other. It ranges from -1 to 1 for completely anti-correlated (one decreases while other increases) and completely correlated (both increase or decrease together) quantities respectively.

Mathematical Formulae-

1) To determine the correlation coefficient between 2 stocks i and j:

$$\rho_{i\,j} = \langle Y_i\, Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle \;/\; \sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}$$

where ;

$\rho_{i\,j}$ is the correlation coefficient between stocks i and j.

$Y_i = \ln(P_i(t)) - \ln(P_i(t-1))$ where $P_i(t)$ is the closing price of stock i at day t.
<x> is the temporal average performed on all the trading days during the investigated period.

2) To determine distance or weight of edge between 2 stocks i and j:

$$d(i,j) = 1 - (\rho_{i\,j})^2$$

Distance is so chosen for it to satisfy all 3 of euclid's axioms which are:

- d(i,j) = 0 if and only if i=j
- d(i,j) = d(j,i)
- d(i,j) = d(i,k) + d(k,j)

Interpretation of graphs in the paper-

1) Figure 1 (a) is the minimum spanning tree connecting the 30 stocks used to compute the Dow Jones industrial average. It has been obtained using calculating the correlation coefficient matrix and the corresponding distance matrix.
   The edges have been colour coded for different weight ranges.
   The stocks which are highly correlated with one another have a "lighter" edge connecting them and those which are not/less correlated have a "heavier" edge connecting them.
   It is observed that stocks of companies belonging to the same industry or sub industry or in other words stocks which are homogeneous with respect to economic activities of the company are found grouped together in the MST and they are connected by comparatively lighter edges.

   Figure 1 (b) is the hierarchical tree of the subdominant ultrametric associated with the MST of the 30 stocks.
   From this tree, the above mentioned group of stocks are easily detected as those with lesser distance between them.
   Some such examples are :

- Oil companies - Exxon, Chevron, Texaco
- Raw material companies - International Paper and Alcoa
- Companies working in the sector of non durable consumer products (Procter and Gamble) and food and drinks (Coca Cola).

2) Figure 2 shows partial regions of the MST of the S&P 500 stocks.
   The edges are similarly colour coded as in figure 1.
   The figure is divided into 4 panels each a subpart of the MST.
   Each panel comprises of stocks homogeneous with respect to economic

activities of the company connected by comparatively “lighter” edges ranging from 0.65-0.75 indicating high correlation among them.
Panel 1 contains financial services companies, capital goods companies, retailing companies, consumer nondurable companies and food and drink companies.
Panel 2 contains International oil companies, oilfield service companies and other energy companies.
Panel 3 contains companies working in the industry sector of electrical utilities and the subindustry sector of telecommunications.
Panel 4 shows a group of companies of the raw material sector. It further has 2 sub groups-

- Companies working in the industry sector of metals and sub industry nonferrous materials.
- Companies working in the industry sector of forest products and packaging.

All 4 panels are connected by dashed edges indicating that the distance between them is greater than 0.81 indicating that the correlation among them is less.

3) Figure 3 is the hierarchical tree of the subdominant ultrametric associated with the MST of the 500 stocks.
Stocks belonging to the same industry, sub industry or sector are indexed together to avoid congestion in the graph and the lesser distance between them is verified in this graph.

Conclusion drawn from the graphs-

Each of the MSTs and hierarchical tree reiterate that stocks belonging to the same industry, sub industry or sector are more likely to be affected similarly by common economic factors specific to that particular group and hence are highly correlated with each other.
Whereas stocks with low correlation are affected by broader common economic factors which are common to all stocks.

Conclusion-

The MST obtained for the stocks using the distance metric and the hierarchical tree of the subdominant ultrametric provides a topological arrangement for the stocks which is able to provide a meaningful economic taxonomy.
This topology is useful in searching for common economic factors that affect the price of specific groups of stocks.
The entire MST as well as the hierarchical tree is obtained solely from the closing price information of the stocks thus indicating that time series of stock prices carry valuable and detectable economic information.