Hierarchical Cluster Analysis on the Hedge Funds

The Rock Creek Group
December 15, 2010

The Rock Creek Group uses agglomerative hierarchical cluster analysis to identify how similar or dissimilar hedge fund managers are from one another. This assists with constructing hedge fund of funds portfolios that are well diversified and appropriate for exploiting market opportunities.

I. Introduction

The cluster analysis is an empirical tool that incorporates a range of mathematical and statistical techniques. The analysis assigns entities, in this particular case hedge funds, to subgroups or clusters based on the similarity of some attribute of the entities. The assignment is made so as to minimize the within-cluster variance of the attribute measure and maximize the across-cluster variance of this measure. The purpose of the cluster analysis is to classify a large group of entities into a small number of sub-groups made up of relatively homogeneous entities, thereby simplifying decision making related to these entities. Cluster analysis is employed in a wide variety of fields, such as taxonomy, genetics, bioinformatics, cosmology, pattern recognition, machine learning, market segmentation, and social networking.

This paper outlines how the Rock Creek Group uses cluster analysis for identifying hedge funds groups that maybe more accurate than the traditional manner of having managers self select the strategy to which they belong.

II. Cluster Analysis

In the cluster analysis, the first and most important step is selecting the attribute on which the clustering is to be done. For an attribute *d,* the distance between hedge fund *i* and *j* is d_{ij}. We use different attributes for clustering, i.e. correlation, mutual information [(1)], country and sector exposures etc. Since the cluster analysis may be used across multiple attributes, i.e. applied in a higher dimensional parameter space, we select Euclidean norm

$$D_{ij} = \sqrt{\sum_d d_{ij}^2}$$

where d_{ij} is the distance between hedge fund i and j for a specific attribute d and D_{ij} is the total distance between hedge fund i and j across all attributes *d.*

Agglomerative hierarchical cluster analysis is a bottoms-up approach. It begins with a single hedge fund as a cluster unto itself. At each successive step smaller clusters are merged into larger clusters. The final step is when all hedge funds in the sample space are agglomerated into one cluster. The merger of clusters at each step is determined by the linkage criterion, which is the fitted pair-wise distance between clusters. There are a number of choices for the linkage criterion. We have chosen the mean linkage criterion for its stability

$$L_{AB} = \frac{1}{|A||B|}\sum_{i\in A}\sum_{j\in B} D_{ij}$$

where A and B are two clusters and L_{AB} is the fitted distance between clusters A and B.

The result of a cluster analysis is usually represented in a tree-structure dendrogram, which is helpful to visualize the relations between entities, i.e. hedge funds.

III. Example

One of the attributes used for grouping hedge funds is the correlation of return series. The distance d_{ij} between hedge funds i and j is

$$d_{ij} = 1 - \rho_{ij} ,$$

where ρ_{ij} is the correlation between the two hedge funds. In this note, we use a portfolio with 16 hedge fund managers as an example. The correlation matrix is displayed in Table 1.

Table 1 : Correlation matrix of an example portfolio with 16 managers is shown below.

	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16
M 1	1.00	0.18	0.47	0.63	0.18	0.36	0.28	0.57	0.53	0.74	0.54	0.25	0.38	0.53	0.47	0.64
M 2	0.18	1.00	0.45	0.21	0.12	0.68	0.06	0.43	0.54	0.16	0.43	0.45	0.36	0.76	0.62	0.47
M 3	0.47	0.45	1.00	0.72	(0.07)	0.66	0.08	0.64	0.85	0.56	0.81	0.61	0.33	0.90	0.87	0.78
M 4	0.63	0.21	0.72	1.00	(0.03)	0.47	0.19	0.68	0.72	0.56	0.71	0.39	0.34	0.67	0.69	0.67
M 5	0.18	0.12	(0.07)	(0.03)	1.00	0.08	0.15	0.03	0.03	0.35	(0.02)	(0.16)	0.34	0.14	(0.06)	0.03
M 6	0.36	0.68	0.66	0.47	0.08	1.00	(0.14)	0.48	0.78	0.40	0.60	0.55	0.32	0.83	0.78	0.60
M 7	0.28	0.06	0.08	0.19	0.15	(0.14)	1.00	0.36	0.12	0.32	0.22	(0.15)	0.29	(0.08)	0.07	0.25
M 8	0.57	0.43	0.64	0.68	0.03	0.48	0.36	1.00	0.72	0.53	0.66	0.32	0.54	0.71	0.65	0.64
M 9	0.53	0.54	0.85	0.72	0.03	0.78	0.12	0.72	1.00	0.55	0.87	0.59	0.42	0.87	0.92	0.84
M 10	0.74	0.16	0.56	0.56	0.35	0.40	0.32	0.53	0.55	1.00	0.57	0.29	0.42	0.65	0.45	0.57
M 11	0.54	0.43	0.81	0.71	(0.02)	0.60	0.22	0.66	0.87	0.57	1.00	0.56	0.35	0.84	0.87	0.94
M 12	0.25	0.45	0.61	0.39	(0.16)	0.55	(0.15)	0.32	0.59	0.29	0.56	1.00	0.03	0.73	0.66	0.46
M 13	0.38	0.36	0.33	0.34	0.34	0.32	0.29	0.54	0.42	0.42	0.35	0.03	1.00	0.55	0.39	0.40
M 14	0.53	0.76	0.90	0.67	0.14	0.83	(0.08)	0.71	0.87	0.65	0.84	0.73	0.55	1.00	0.86	0.79
M 15	0.47	0.62	0.87	0.69	(0.06)	0.78	0.07	0.65	0.92	0.45	0.87	0.66	0.39	0.86	1.00	0.84
M 16	0.64	0.47	0.78	0.67	0.03	0.60	0.25	0.64	0.84	0.57	0.94	0.46	0.40	0.79	0.84	1.00

Although correlation is an important diversification measure, a correlation matrix often contains too many numbers making it difficult to interpret and incorporate in asset allocation. The correlation matrix in Table 1 has $\frac{16\times15}{2} = 120$ pair-wise correlations. A typical fund of hedge funds portfolio will have many more managers and as a result many more pair wise correlations. We use cluster analysis to make the data more tractable for decision making. The dendrogram from the cluster analysis is shown in Figure 1.

The cluster analysis can help with asset allocation in multiple ways.

- It can identify managers with the highest correlations to one another, i.e. manager 11 and manager 16. Those managers may be substitutes for one another if necessary.
- It can also identify groups of managers with similar return patterns, i.e. the core cluster shown in Figure 1.
- It can help with increasing the overall diversification of the portfolio by identifying the diversification cluster shown in Figure 1.

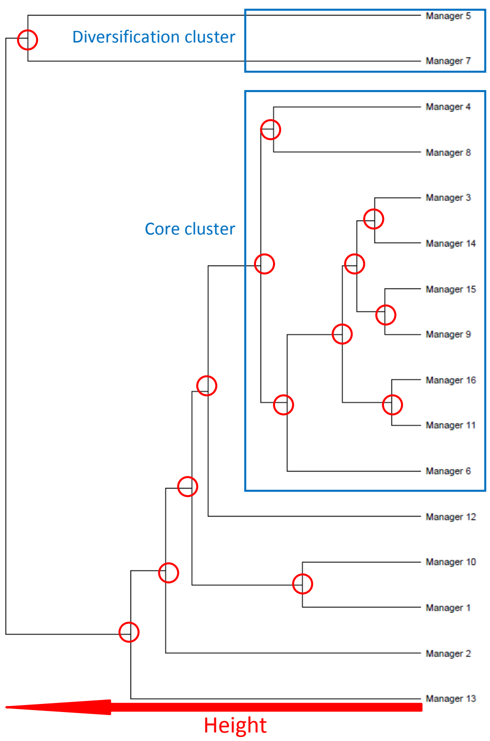

Figure 1: Dendrogram of clusters in the example portfolio. The red circles indicate the node in the dendrogram. The distance between two managers or clusters is the height of their lowest common node.

IV. Cluster Analysis on Other Attributes

The cluster analysis may be applied to attributes other than the correlation of return series. For example it may be applied to:

- RCG mutual information correlation [1]. The RCG mutual information correlation is a correlation measure that is distinct from the Pearson or Spearman correlation measures. It is based on the concept of information entropy. It is able to indentify non-linear relationships and is not affected as much by outliers. The result of this cluster analysis on this attribute for the portfolio of 16 hedge fund managers is shown in Figure 2. The main structure of the result is quite similar to the cluster based on the more standard correlation measure.

- Exposure data. The exposure cluster is a forward-looking analysis, as compared to the return correlation analysis that is more backward looking. It identifies groups of managers who may be expected to have similar returns based on their current exposure.

- Range of risk and performance measures, such as volatility, extreme expected loss/gain, marginal utility, Sharpe ratio, and RCG Performance ratio. If used in combination, simple Euclidean norms require that these measures are orthogonal to one another. Our research has identified several orthogonal risk and performance measures [2].

Quantitative cluster analysis should be combined with the qualitative assessments to be useful for decision making. For example, hedge fund managers should be qualitatively categorized into groups according to characteristics, such as strategy, trading style, and organizational characteristics of the managers. This qualitative grouping should complement the quantitative cluster analysis to identify suitable manager clusters for portfolio construction.

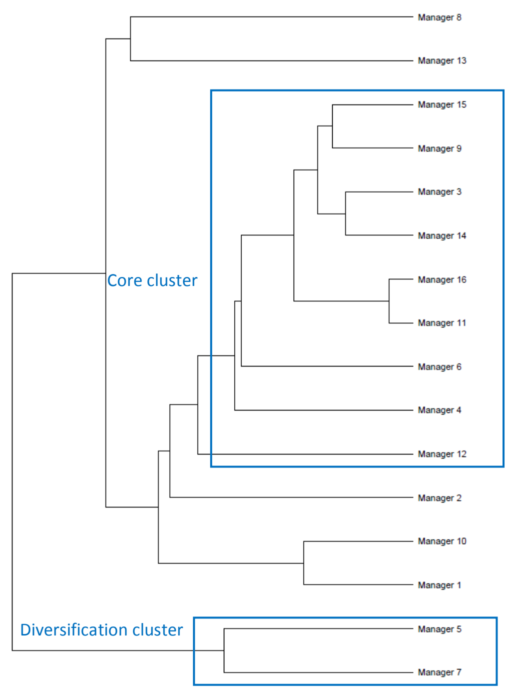

Figure 2 : Dendrogram of mutual information clusters in the example portfolio.

V Modularity of Clusters

Cluster modularity measures the degree to which the individual entities (or managers) in a portfolio are separate and distinct from one another. In general, the greater the modularity of a cluster, the higher the similarity within a cluster and lower the similarity across clusters. Increased modularity leads to better diversification and greater ability to withstand external shocks to the portfolio. The modularity measure is used in many fields of physical science, and has been recently used to study world trade network [(3)].

In a hierarchical cluster, modularity is quantified by the cophenetic correlation coefficient (CCC):

$$CCC = \frac{\sum_{i<j}(D_{ij} - \bar{D})(L_{ij} - \bar{L})}{\sqrt{\sum_{i<j}(D_{ij} - \bar{D})^2 \sum_{i<j}(L_{ij} - \bar{L})^2}},$$

where D_{ij} is the distance between hedge fund i and j, $\bar{D}$ is the average distance in the portfolio, L_{ij} is the linkage between hedge fund i and j, $\bar{L}$ is the average linkage in the cluster structure.

We evaluated 18 distinct hedge fund portfolios. The average correlation between the hedge fund managers in each portfolio is about the same, ~ 0.4. However, their modularity was quite different. Figure 3 shows the modularity of the 18 portfolios as of April 2010 and their respective returns in May 2010 in the context of a relatively large market shock.

.

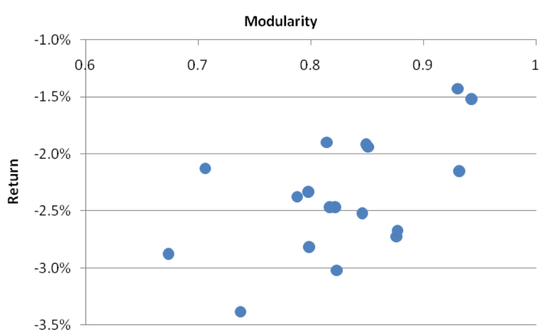

Figure 3 : Modularity vs. Return for 18 portfolios.

There is a strong relationship between portfolio returns and its modularity. The portfolios with greater modularity withstand the market shock better, even though the average correlation between the managers is about the same.

VI. Conclusion

The cluster analysis is a useful tool in the asset allocation process. It captures the internal structure of correlations between managers. It helps identify similar managers, discover distinct groups, and achieve portfolio diversification. This analysis can be undertaken using different quantitative attributes and suitably complemented with qualitative measures. Furthermore, the modularity of the portfolio cluster is a useful measure of the portfolio's ability to withstand a market shock.

Bibliography

The Rock Creek Group. *Mutual Information Correlation.* 2009.

The Rock Creek Group. *Advanced Risk and Performance Measures.* 2009.

He, Jiankui and Deem, Michael W. *Structure and Response in the World Trade Network.* 2010.