HIERARCHICAL CLUSTERING: PREDICTION OF SYSTEMATIC UNDERPERFORMANCE

The popularity of machine learning methods has increased immensely in recent years. One such method is hierarchical clustering; an unsupervised learning algorithm that uses the distance between data points, and assigns them into clusters with similar traits.

In this article, hierarchical clustering will be applied to funds in order to group them if they demonstrate a similar statistical behaviour in returns. By testing the hypothesis that there are differences in average return between the assets belonging to the same group, it will be evaluated whether it is possible to predict such differences in future returns as well.

Thereafter, an additional analysis is performed. It will be evaluated whether the fund fee can be used as a predictor of relative performance between funds. This will answer a question of substantial importance - will funds with higher fees on average outperform similar funds with lower fees?

ASSET CLUSTERING

Numerous studies have addressed the challenges of constructing an optimal portfolio from a set of assets. One of such difficulties is that the in-sample optimal portfolios are often outperformed out-of-sample by simple strategies such as an equally weighted portfolio.

This is where hierarchical clustering becomes useful. The assets are clustered such that each group only contains assets that are assumed to behave equally, both regarding joint- and marginal distributional properties. Consequently, the problem of judging an asset's performance relative to its addition to the optimal portfolio's distributional properties can be avoided. The challenge is simply reduced to finding systematic differences of asset performance within each group.

Clustering algorithms consist of two parts – defining a distance measure, and defining a method to construct clusters based on that distance. Marti *et al.* (2016) proposes a linear combination of the Hellinger distance, measuring the distance regarding marginal distributional properties, and the rank correlation Spearman's rho, to measure the distance between assets.

These will also be our distance measures of choice. However, instead of defining the distance measure as a combination of the two, the clustering will be performed in two steps; the first round clusters the assets based on their joint distributional properties, while the second round clusters with regard to their marginal distributional properties. This makes it possible to define a maximum distance allowed within a cluster for each of the concerned distances.[1]

Having defined the distance measures, only the actual clustering algorithm remains. Agglomerative clustering was used to ensure that the most similar assets end up in the same group. As linkage method "average group distance" was chosen, as "complete-linkage" was found to produce too large clusters when working with many clusters. The algorithm for clustering the assets is described below:

1. Cluster the assets into two groups using the correlation distance measure.
2. For each group:
   - ◇ If the maximum distance within the group is smaller than some threshold: Stop.
   - ◇ Else: Use the group as input to step 1.
3. Use every cluster produced using the correlation distance measure as input to step 1, but change the distance measure to the Hellinger distance.

HYPOTHESIS TESTING

Since all assets in one cluster are assumed to have the same joint- and marginal distributional properties, it is possible to test systematic under/over-performance via hypothesis testing of the average return. The null hypothesis states that the average return, $\overline{r}_A$, of one asset is the same as the average return of another asset, $\overline{r}_B$, while the alternative hypothesis states that the average return of asset A is less than that of asset B:

$$\begin{aligned} H_0 &: \overline{r}_A = \overline{r}_B, \\ H_1 &: \overline{r}_A < \overline{r}_B \end{aligned} \tag{1}$$

[1] This would of course have been possible using the combined distance measure as well, but by splitting the clustering into two rounds we avoid the trade-off between joint- and marginal distributional properties.

It is then possible to reject H_0 on significance level α if the resulting p-value is smaller than α:

$$P(r_B - r_A \geq \overline{r}_B - \overline{r}_A | H_0) < \alpha \tag{2}$$

Under the assumption that all differences are independent and identically distributed this probability can be calculated analytically using the central limit theorem. However, to avoid making unnecessary assumptions, and relying on the central limit theorem for a possibly small number of observations, the probability is instead calculated using a non-parametric bootstrap approach called a permutation test.

HYPOTHESIS TESTING FOR A CLUSTER OF ASSETS

If a cluster consists of n assets, there will for every asset be $n-1$ possible p-values. However, it is sufficient that an asset is dominated by only one other asset. This implies that the p-value of interest is the smallest p-value. To further motivate this, consider an asset A that outperforms all other assets but one in a cluster. If there is a systematic difference in performance in favour of the remaining asset, it would still never be favourable to invest in asset A.

Hence, the hypothesis that we would like to test is:

$$\begin{aligned} H_0 &: \overline{r}_A \geq \overline{r}_j, \ for \ all \ j = 1, .., n, \ j \neq A, \\ H_1 &: \overline{r}_A < \overline{r}_j, \ for \ any \ j = 1, .., n, \ j \neq A, \end{aligned} \tag{3}$$

where H_0 is rejected on significance level α if:

$$P\left(min\left[P_1, P_2, ..., P_{n-1}\right] < min\left[P_{observed}\right] | H_0\right) < \alpha, \tag{4}$$

where P_i denotes the p-value corresponding to the test of asset i versus asset A, and $min(P_{observed})$ is the smallest observed p-value.

Assuming the null hypothesis is true, the p-values are uniformly distributed between zero and one. If we additionally assume that the p-values are independent it is possible to calculate the probability:

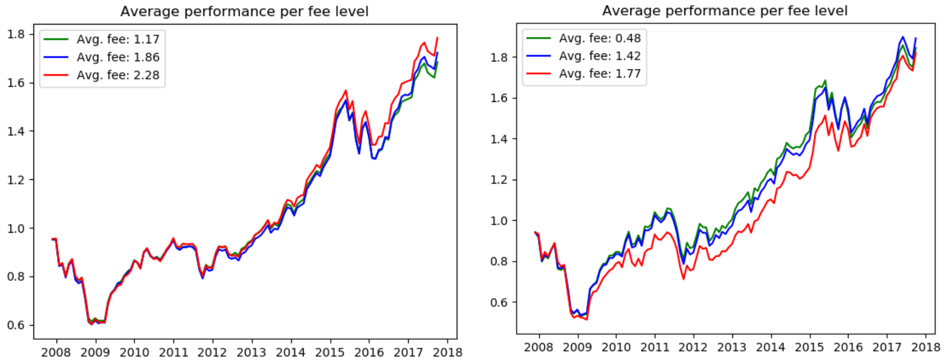
$$\begin{aligned} & P\left(min\left[P_1, P_2, ..., P_{n-1}\right] < min\left[P_{observed}\right] | H_0\right) \\ & = 1 - P\left(\bigcap_{i=1}^{n} P_i \geq min\left[P_{observed}\right] | H_0\right) = \\ & = 1 - \left(1 - min\left[P_{observed}\right]\right)^{n-1}. \end{aligned} \tag{5}$$

If instead the p-values would be completely dependant, all p-values would be exactly the same and no correction for the number of assets in the cluster would be required. As a result, a lower and upper limit is obtained for the p-value of the hypothesis stating that there is at least one asset with a higher average return than the asset of interest. If the upper limit of the interval is below the significance level α, it is certain that H_0 can be rejected on that level. However, if α is in the middle of the interval it is not certain whether or not H_0 can be rejected. Lastly, if the lower limit of the interval of the p-value is above the significance level α we can be certain that the null hypothesis can not be rejected based on this analysis.

FEES

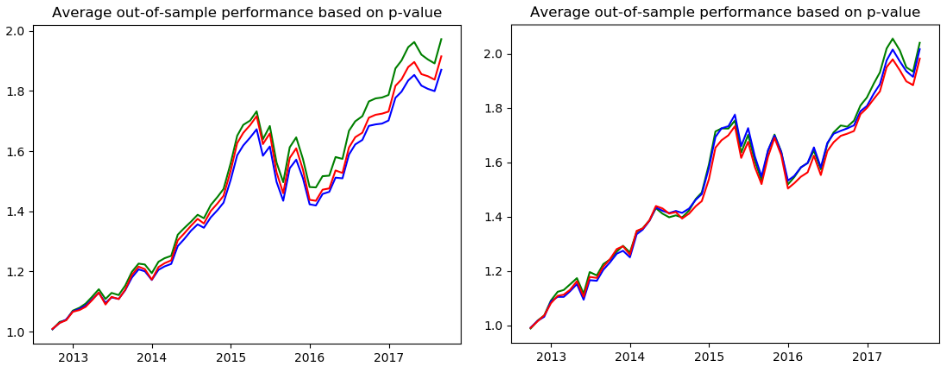
In an efficient market, there is no room for different fees within a cluster. A higher fee would imply systematic under-performance. However, as market efficiency is always a hot topic for debates, there will be an additional evaluation analysing whether the fee of the assets can result in a significant difference in average return. By grouping the assets in each cluster into three equally sized groups based on their fee, and comparing the average returns of the groups using the methodology described in the previous section, it is possible to calculate the significance of the fees' impact on the average return. The comparison will be performed both within each cluster, and at an all asset-level. Since the fee is not calibrated to the asset returns, the analysis can be performed in-sample.

RESULTS



(a) (b)

Figure 1: *Average performance per fee level for (a) all funds and (b) Swedish funds. The p-value corresponding to the null hypothesis that the group with the highest fee has the same average return as the group with the lowest fee is 0.86 for all assets and 0.47 for the Swedish funds.*



(a) (b)

Figure 2: *Average performance based on p-value for (a) all funds and (b) Swedish funds. The p-value corresponding to the null hypothesis that the group with the lowest p-value has the same average return as the group with the highest p-value is 0.09 for all assets and 0.26 for the Swedish funds.*

The analysis was conducted on real data from 384 funds within different geographical and industrial sectors. The joint results for all assets and the results for a cluster consisting of Swedish funds are exhibited in the plots above. To form the joint results for all assets, the funds in each group were split into three subgroups based on their fees and p-values, respectively. Then, all assets belonging to, for instance, the groups with the highest fees were merged to form the high fee group for all assets. The lines seen in

Figure 1 and 2 represent the performance of an equally weighted portfolio of all assets in such a subgroup based on fee or p-value. While most results were not statistically significant, the results trend towards that the p-value could act as a predictor of future relative performance between funds. The fees, on the other hand, provided no such indications; neither a high nor a low relative fee seems to be a significant predictor of future performance.

Bibliography

Marti, G., Nielsen, F., Donnat, P., & Andler, S. 2016. On clustering fi nancial time series: a need for distances between dependent random variables.

CONTACT US

Do you want to know more about Kidbrooke Advisory?
Please contact us for more information.

e. info@kidbrooke-advisory.com

a. Kidbrooke Advisory AB
Engelbrektsgatan 7
SE-114 32 Stockholm
Sweden

SANNA BRANDEL
Analyst
+46(0)732 626 675
sanna.brandel@kidbrooke.com

FILIP PERSSON MÖRK
Associate
+46(0)709 393 086
filip.mork@kidbrooke.com

DISCLAIMER