Asset Pricing & Portfolio Management

Hierarchical Risk Parity Portfolios

Jorge Miguel Bravo, PhD

NOVA IMS

October 4, 2023

Agenda

1. Introduction
2. Hierarchical Risk Parity Portfolios (HRPP)
   - Hierarchical Tree Clustering
   - Linkage criteria
   - Matrix Seriation
   - Recursive Bisection
3. Final remarks

- **Readings**:
  - Lopez de Prado, M., 2016, Building Diversified Portfolios that Outperform Out-of-Sample, Journal of Portfolio Management, 2016. Available at SSRN: https://ssrn.com/abstract_id=2708678
  - Maillard, S., Roncalli, T. and Teiletche, J., 2008, On the properties of equally-weighted risk contributions portfolios, Working Paper, Available at SSRN: https://ssrn.com/abstract_id=1271972

Background

- Lopez de Prado (2016) introduced the Hierarchical Risk Parity (HRP) approach to address three major concerns of quadratic optimizers: instability, concentration, and underperformance
  - The instability problem of traditional allocation strategies is mitigated in the HRP approach, as it does not rely on the estimation of the expected returns or require the inversion of the covariance matrix
  - The concentration problem concerns mainly the minimum-variance allocation strategy, which drives extreme concentration because of its goal of minimizing the portfolio's variance
  - The underperformance issue may refer to many portfolio optimization methods failing to outperform the simple EWP ($1/N$) portfolio consistently (DeMiguel et al., 2009)
- Unlike the traditional asset allocation strategies, HRP incorporates the notion of hierarchy
- Instead of viewing all assets in a portfolio as substitutes, an investor would undoubtedly consider certain assets as substitutes and others as complementary

What is Hierarchical Risk Parity (HRP)?

- HRP is a portfolio optimization technique consisting of three steps:
  - **Hierarchical Tree Clustering**
    - Group similar investments into clusters, based on a proper distance metric and using single linkage
    - The method takes advantage of the relationship among financial assets (correlation) to create a hierarchical structure that can be plotted as a dendrogram
    - A dendrogram is a diagram representing a tree that shows the hierarchical relationship between objects
  - **Matrix Seriation**
    - the assets in the dendrogram are sorted minimizing the distance between leafs, Lopez de Prado called this process quasi-diagonalization
    - Reorganize the rows and columns of the covariance matrix, so that the largest values lie along the diagonal
  - **Recursive Bisection**
    - Split the weights along the dendrogram using naive risk parity (weights based on the inverse of asset's risk) from the top of the tree to the leafs

Tree Clustering

- In the first stage of the HRP approach, a hierarchical clustering algorithm is used to group "similar" assets
- This requires to decide what metric to use for similarity
- Lopez de Prado (2016) uses the Pearson correlation coefficient as the root of similarity between the assets
- He starts with the correlation matrix and converts it into a distance matrix

$$\mathbf{D} = \{d_{i,j}\}_{i,j=1,\ldots,n} = \begin{bmatrix} 0 & d_{1,2} & \cdots & d_{1,n} \\ d_{2,1} & 0 & \cdots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n,1} & d_{n,2} & \cdots & 0 \end{bmatrix}$$

where

$$d_{i,j} = \sqrt{\frac{1}{2}\left(1 - \rho_{i,j}\right)}$$

- For perfectly correlated assets, $\rho_{i,j} = 1$, the distance is zero. Conversely, perfectly negatively correlated assets, $\rho_{i,j} = -1$, have a distance of one

Tree Clustering

- This distance only considers assets as pairs
- Yet it would be advantageous to also consider the role that each asset plays in the context of the asset universe
- Imagine two assets with low correlation, but with similar relationship with the rest of the assets

Example

Assets A and B have low correlation, but the correlation between A and C is similar to the correlation between B and C. Then, it would be beneficial to consider these assets more similar than initially thought

- The HRP takes the Euclidean distance between the columns of the original distance matrix $\mathbf{D}$ and forms the Euclidean distance matrix $\tilde{\mathbf{D}}$:

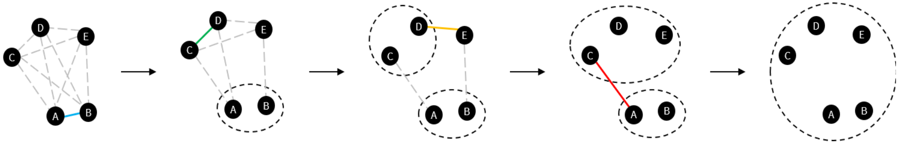
$$\tilde{\mathbf{D}} = \{\tilde{d}_{i,j}\}_{i,j=1,\ldots,N}$$

where

$$\tilde{d}_{i,j} = \sqrt{\sum_{n=1}^{N} \left(d_{n,i} - d_{n,j}\right)^2}$$

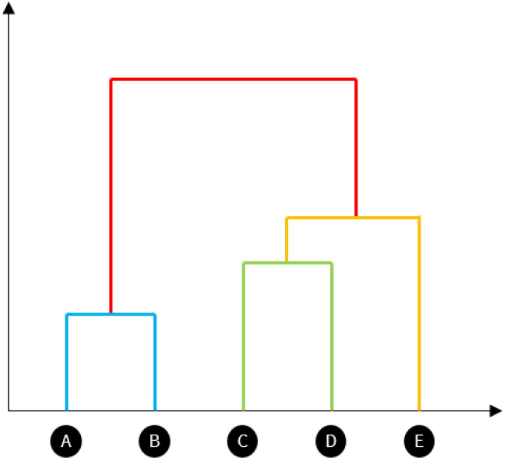
Tree Clustering

- The difference between the $d_{i,j}$ and $\tilde{d}_{i,j}$ is subtle, where the first measures the distance of the column vectors of the correlation matrix, and where the second measures the distance on the column vectors of $\mathbf{D}$, that is, a distance of distances
- The next step is to perform the clustering based on a linkage criteria clustering algorithm (originally single linkage)
  - Single Linkage defines the distance between two clusters as the shortest possible distance between them



Tree Clustering

Additionally, a dendrogram can be constructed to illustrate the notion of hierarchical clustering



In the dendrogram, the height of the dendrogram indicates the order in which the clusters were joined

Toy example

- **Correlation matrix $\boldsymbol{C}$:**

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0.7 & 0.2 \\ 0.7 & 1 & -0.2 \\ 0.2 & -0.2 & 1 \end{bmatrix}$$

- **Correlation-based distance matrix $\boldsymbol{D}$:**

$$\boldsymbol{D} = \begin{bmatrix} 0 & 0.3873 & 0.6325 \\ 0.3873 & 0 & 0.7746 \\ 0.6325 & 0.7746 & 0 \end{bmatrix}$$

- **Euclidean distance matrix of correlation distances $\tilde{\boldsymbol{D}}$:**

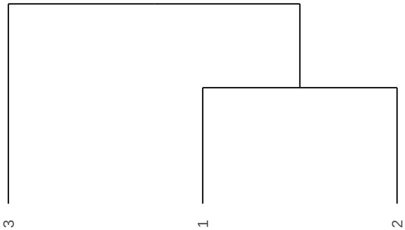
$$\tilde{\boldsymbol{D}} = \begin{bmatrix} 0 & 0.5659 & 0.9747 \\ 0.5659 & 0 & 1.1225 \\ 0.9747 & 1.1225 & 0 \end{bmatrix}.$$

Hierarchical clustering and dendrograms

Consider a toy example with distance matrix:

$$\tilde{\boldsymbol{D}} = \begin{bmatrix} 0 & 0.5659 & 0.9747 \\ 0.5659 & 0 & 1.1225 \\ 0.9747 & 1.1225 & 0 \end{bmatrix}$$

The dendrogram groups first the first and second elements since they have the smallest distance:



Tree Clustering

- The clustering process can be stored in a linkage matrix Y. Taking the previous example, the linkage matrix will have four columns and $(N-1)$ rows, where each row m represents one cluster and each column holds the following information:

$$Y = (y_{m_1}, y_{m_2}, y_{m_3}, y_{m_4})$$

  - y_{m_1} = reference to the left constituent of cluster m
  - y_{m2} = reference to the right constituent of cluster m
  - y_{m_3} = the distance between the two constituents of cluster m
  - y_{m_4} = the number of assets in cluster m

m	y_{m_1}	y_{m_2}	y_{m_3}	y_{m_4}
1	A	B	d_1	2
2	C	D	d_2	2
3	CD	E	d_3	3
4	AB	CDE	d_4	5

Tree Clustering: summary

Algorithm 1 Hierarchical Clustering using HRP

1. Estimate $N \times N$ correlation matrix $\boldsymbol{\rho} = \{\rho_{i,j}\}_{i,j=1,\ldots,N}$
2. Convert the correlation matrix into a distance matrix $\mathbf{D} = \{d\}_{i,j=1,\ldots,N}$ where $d_{i,j} = \sqrt{\frac{1}{2}(1-\rho_{i,j})}$
3. Compute a Euclidean distance matrix $\tilde{\mathbf{D}} = \{\tilde{d}_{i,j}\}_{i,j=1,\ldots,N}$ where $\tilde{d}_{i,j} = \tilde{d}(X_i, X_j) = \sqrt{\sum_{n=1}^{N}(d_{n,i} - d_{n,j})^2}$
4. Let us denote the set of clusters as U:
   (a) Construct the first cluster (i^*, j^*) as $U[1] = \underset{i,j}{\operatorname{argmin}}\ \tilde{D}(i,j)$
   (b) Using a single linkage, update the distance matrix by calculating the pairwise distances of the newly formed cluster and other items
   (c) Continue combining assets recursively until we are left with only one single cluster
5. Summarize the information in linkage matrix $\mathbf{Y} \in \mathbb{R}^{(N-1)\times 4}$

Tree Clustering

Linkage Criteria

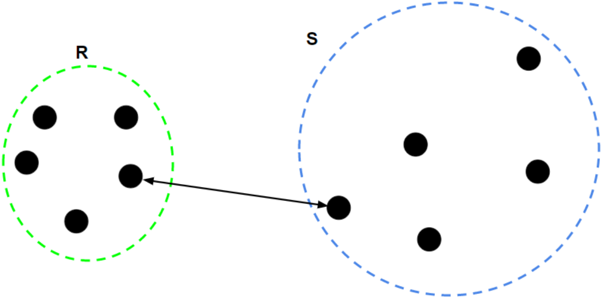
- The linkage criterion determines the distance between clusters
- There are four primary linkage criteria:
  - Single linkage
  - Complete linkage
  - Average linkage
  - Ward's method

Tree Clustering

Linkage Criteria

- **Single Linkage**
  - In single linkage, also called the nearest-neighbor clustering, the distance between two clusters is the minimum distance between members of the two clusters
  - For two clusters R and S, the single linkage returns the minimum distance between two points i and j such that i belongs to R and j belongs to S

$$L(R, S) = \min\left(d_{i,j}\right), \; i \in R, \; j \in S \tag{1}$$

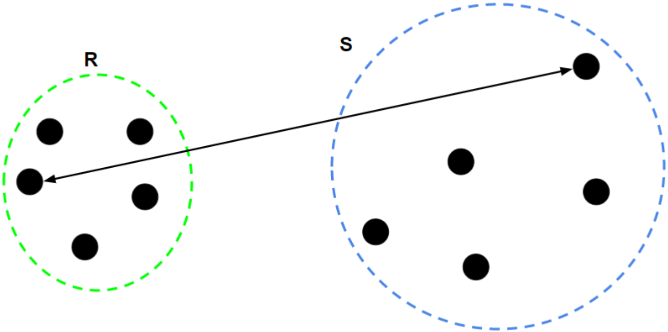


Tree Clustering

Linkage Criteria

- **Complete Linkage**
  - For two clusters R and S, the single linkage returns the maximum distance between two points i and j such that i belongs to R and j belongs to S

$$L(R,S) = \max\left(d_{i,j}\right), \ \ i \in R, \ j \in S \tag{2}$$



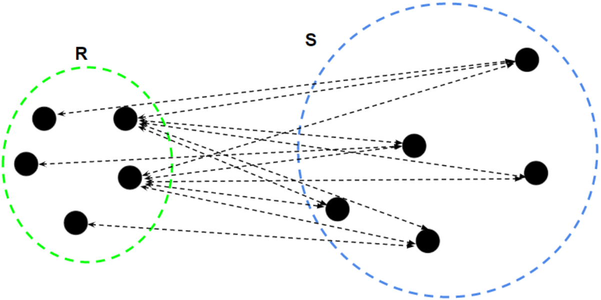
Tree Clustering

Linkage Criteria

- **Average Linkage**
  - For two clusters R and S, first for the distance between any data-point i in R and any data-point j in S and then the arithmetic mean of these distances are calculated. Average Linkage returns this value of the arithmetic mean.

$$L(R,S) = \frac{1}{n_R + n_S}\left[\sum_{i=1}^{n_R}\sum_{j=1}^{n_S} d_{i,j}\right], \quad i \in R,\ j \in S \tag{3}$$

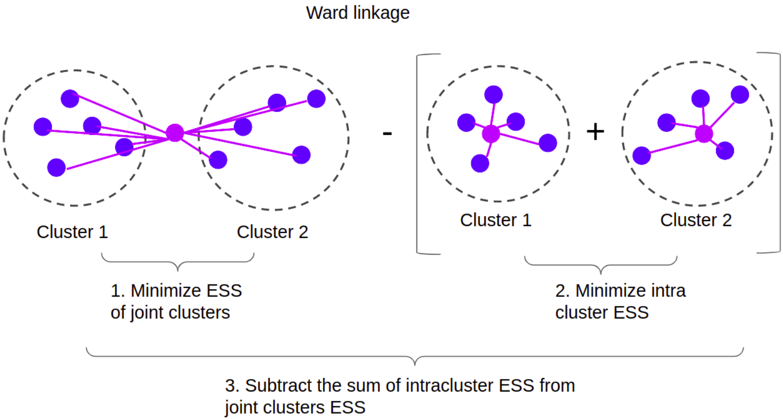
where n_R and n_S denote the number of data-points in R and S



Tree Clustering

Linkage Criteria: Ward's Method

- Also called the minimum variance method, is an agglomerative clustering method proposed by Joe H. Ward Jr. in 1963
- Assumes that a cluster is represented by its centroid (centre point of the cluster), and the proximity between two clusters is measured in terms of the increase in the sum of squares error (ESS) resulting from cluster merging



Matrix Seriation

- Matrix seriation is a statistical method used to reorganize rows or columns of a matrix to enumerate them in an appropriate order
- In the second stage of the HRP algorithm, the rows and columns of the covariance matrix are reorganized so that the largest values lie along the diagonal
- This quasi-diagonalization of the covariance matrix is based on the information from the clustering algorithm in the previous stage of the HRP approach
- This way, the assets in the covariance matrix which are similar will be placed together, and dissimilar assets will be placed far apart

Matrix Seriation

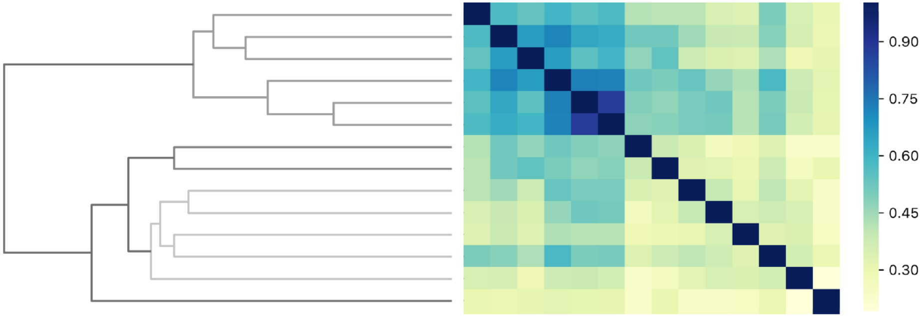


Figure: Illustration of the seriation of a covariance matrix using hierarchical clustering

Recursive Bisection

- In the final stage of the HRP algorithm, the recursive bisection algorithm allocates the portfolio weights
    1. First, the algorithm is initialized by making a list, L, of items, L_0, and assigning a unit weight of one, $\{w_i = 1\}_{i=1,\dots,n}$
    2. Then in a tree-like manner, the algorithm goes from top to bottom and allocates weight each time the clusters splits
    3. The algorithm assumes the tree to be binary, where clusters are recursively split into two equally-sized sub-clusters
    4. The algorithm uses a top-down inverse-variance allocation to determine the final weights of each asset, the weight of each asset n is determined by the intra-cluster inverse-variance allocation

$$w_n = \frac{1/\sigma_n^2}{\sum_{i=1}^{N} 1/\sigma_i^2}$$

5. Finally, the relative weights of each sub-cluster are updated using a split factor, α_i, which is calculated using the inverse-variance allocation between the two clusters

Recursive Bisection

Example

For a portfolio of five assets and $\alpha_i = 0.6$, the list of weights would be updated in the following manner

$$\begin{aligned}
[1, 1, 1, 1, 1] &\longrightarrow [0.6, 0.6, 0.4, 0.4, 0.4] \\
&\longrightarrow [0.36, 0.24, 0.4, 0.4, 0.4] \\
&\longrightarrow [0.36, 0.24, 0.24, 0.16, 0.16] \\
&\longrightarrow [0.36, 0.24, 0.24, 0.096, 0.064]
\end{aligned}$$

Hierarchical Risk Parity Portfolios

Thank you for your attention

Jorge Miguel Bravo
jbravo@novaims.unl.pt
Universidade Nova de Lisboa, NOVA IMS & MagIC
Université Paris-Dauphine PSL
CEFAGE-UE & BRU-ISCTE-IUL
Scopus ID: 57193745638
ORCID: 0000-0002-7389-5103