Portfolio Optimization: Hierarchical Risk Parity

Source: Marco López de Prado, Advances in Financial Machine Learning, Wiley, 2018

Key Points

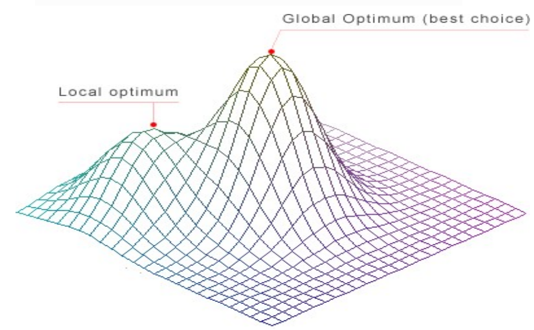
- <u>The problem</u>: Mean-Variance (MV) portfolios are optimal *in-sample*, however they tend to perform poorly *out-of-sample* (even worse than the 1/N naïve portfolio!)
- <u>Two major causes</u>:
  1. Returns can rarely be forecasted with sufficient accuracy
  2. Quadratic optimizers require the inversion of a positive-definite covariance matrix
- <u>A partial solution</u>: To deal with the first cause, some modern approaches drop returns forecasts, e.g. Risk Parity (RP)
- Still, matrix inversion is a major reason why MV and RP underperform out-of-sample (OOS)
- What portfolio optimization method can improve OOS performance?

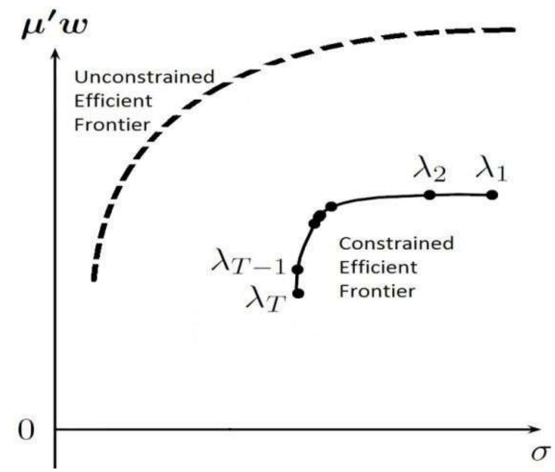
Quadratic Optimization

- In 1956, Harry Markowitz developed the Critical Line Algorithm (CLA):
  - CLA is a quadratic optimization procedure specifically designed for inequality constraints to solve the mean variance portfolio optimization problem: each weight in the allocation has an upper and a lower bound.
  - It finds the exact solution after a known number of steps
  - It ingeniously circumvents the Karush-Kuhn-Tucker conditions through the notion of “turning point”
- A turning point occurs when a previously free weight (not constrained by any bounds) hits a boundary
- The constrained efficient frontier between two neighbor turning points can be reformulated as an unconstrained problem
- CLA solves the optimization problem by finding *the sequence of turning points*

Open source CLA Python library: [Bailey and López de Prado [2013]]()

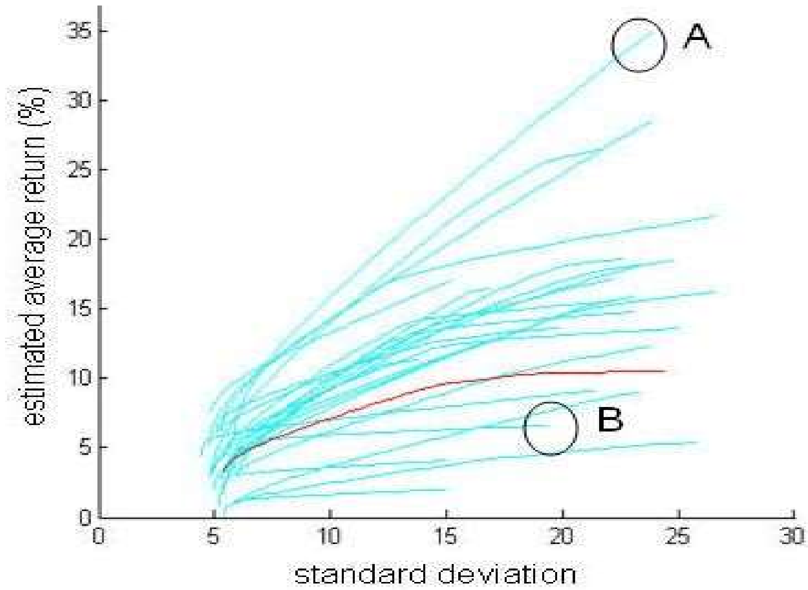
$$
\begin{aligned}
\min_{\mathbf{w}} \quad & w^T \sum w \\
\text{s.t.} \quad & \sum_{i=1}^{n} w_i = 1 \\
& w_i <= u_i \\
& w_i >= l_i
\end{aligned}
$$





Risk Parity

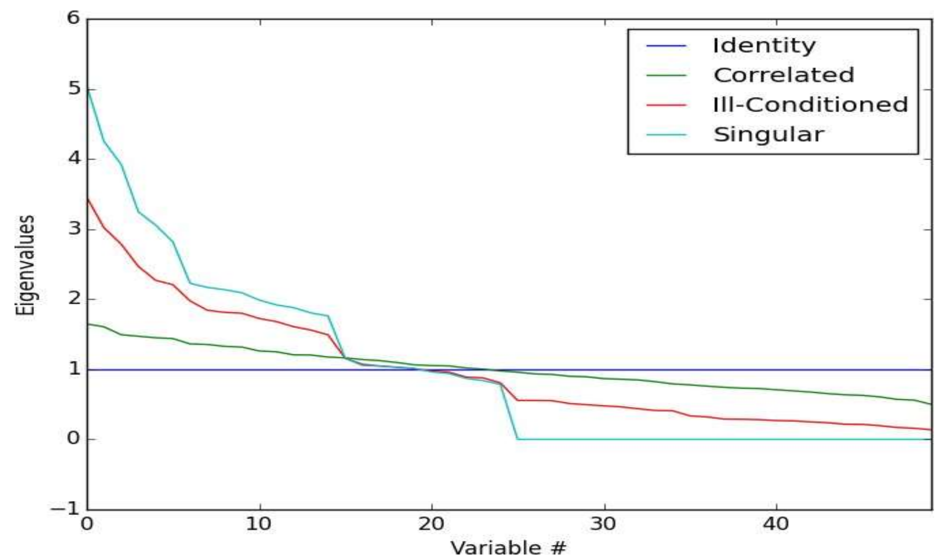
- Numerous studies show that quadratic optimizers in general produce unreliable solutions, e.g. Michaud [1998]
- One major reason for this is that returns can rarely be forecasted with sufficient confidence
- Small forecasting errors can lead to dramatically different Efficient Frontiers



As a consequence, many authors have dropped returns forecasting altogether, giving rise to risk-based asset allocation approaches. E.g.: Risk parity: allocate portfolio based on risk allowing long and short positions.

Markowitz's curse

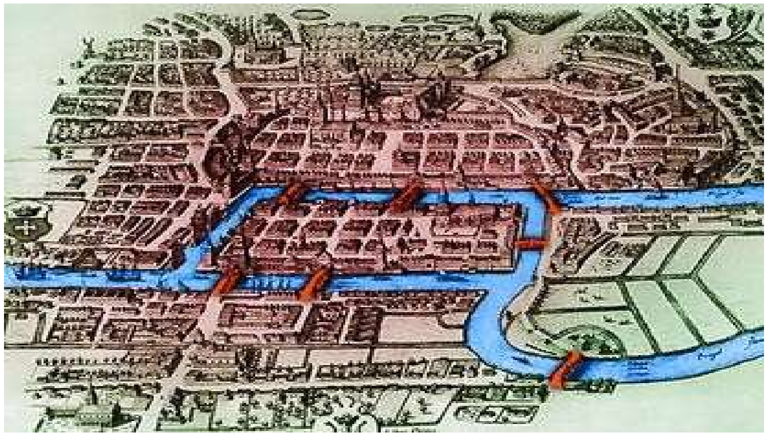
- De Miguel et al. [2009] show that many of the best known quadratic optimizers underperform the Naïve 1/N allocation OOS, even after dropping forecasted returns!
- The reason is, quadratic optimizers require the inversion of a positive-definite covariance matrix
- The *condition number* of a covariance matrix is the ratio between its highest and smallest (in moduli) eigenvalues
- The more correlated the assets, the higher the condition number, and the more unstable is the inverse matrix
- **Markowitz's curse: Quadratic optimization is likely to fail precisely when there is a greater need for finding a diversified portfolio**



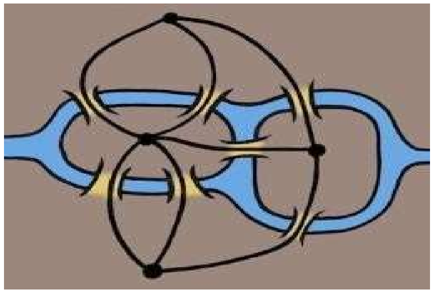
From Geometry To Topology

Topology

Is it possible to walk through the city of Königsberg crossing each bridge once and only once, ending at the starting point?



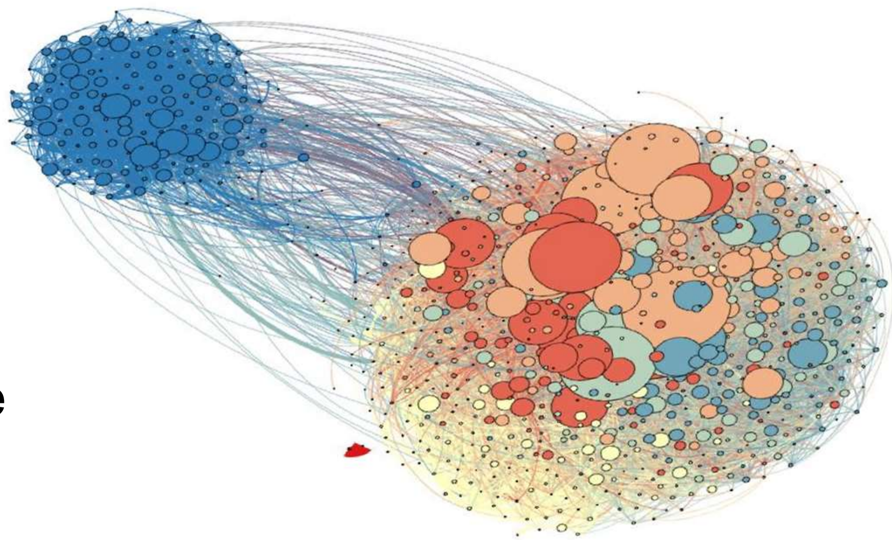
- Around 1735, Leonhard Euler asked this question
- Euler was one of the first to recognize that Geometry could not solve this problem



- **Hint: The relevant information is not the geometric location of the bridges, but their logical relations**

Graph Theory

- A graph can be understood as a **relational map** between pairs of items
- Once considered a branch of Topology, Graph Theory has grown to become a Mathematical subject in its own right
- Graph theory can answer questions regarding the logical structure, architecture and dynamics of a complex system



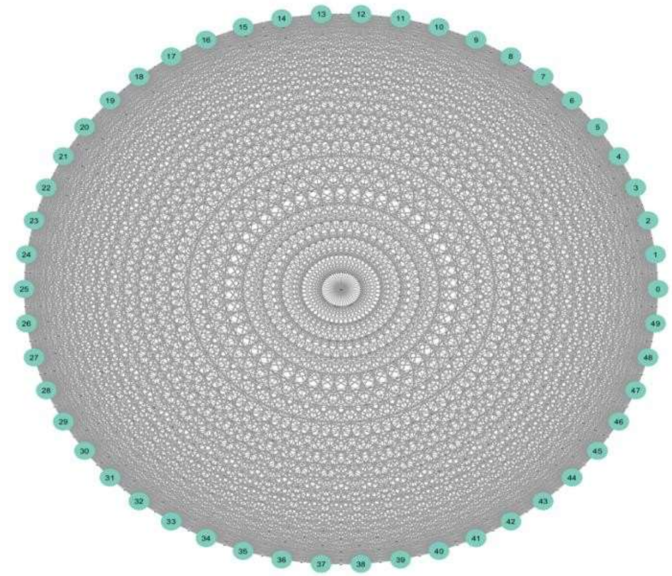
Graph Theory is applied by Google to rank hyperlinks, by GPS systems to find your shortest path home, by LinkedIn to suggest connections, by the NSA to track terrorists... In the example above, Graph Theory is used to derive the political inclination of a community, based on the speeches they re-tweet

Topology & Graph Theory atwork

- Subway plans are not geographic maps, but topological representations of how lines and stations are interconnected
- Adding to that map geographic details would make it more difficult to solve problems such as how to find alternative routes, minimize waiting time, avoid congestion paths, etc.

- Lesson: finance or economic problems can be solved with their topological representations that show only logical and relevant relations between their parts.

What does it mean "inverting the matrix"?

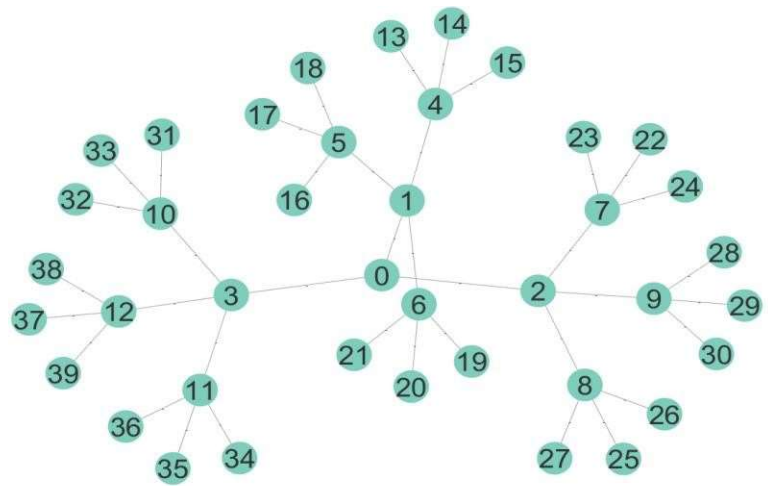


- One reason for the instability of quadratic optimizers is that the vector space is modelled as a complete (fully connected) graph, where every node is a potential candidate to substitute another
- In algorithmic terms, inverting the matrix means evaluating the rates of substitution across the complete graph

- For a numerically-ill conditioned covariance matrix, small estimation errors over several edges lead to grossly incorrect inversions
- Correlation matrices lack the notion of hierarchy, because all investments are potential substitutes to each other
- Intuitively it would be desirable to drop unnecessary edges

Hierarchical Risk Parity (HRP)

- A tree structure introduces two desirable features for portfolio optimization:
  - It has only *N-1* edges to connect *N* nodes, so the weights only rebalance among peers at various hierarchical levels
  - The weights are distributed top-down, consistent with how asset managers build their portfolios, e.g. from asset class to sectors to individual securities.



- The HRP algorithm works in three stages:
  1. **Tree Clustering**: Using hierarchical clustering, group similar investments into clusters, based on a distance metric that depends of the correlation between assets
  2. **Quasi-diagonalization or matrix seriation:** reorganize the covariance matrix using the results of hierarchical cluster so that similar stocks are together and distant from dissimilar stocks.
  3. **Recursive bisection**: Split allocations through recursive bisection of the reordered covariance matrix.

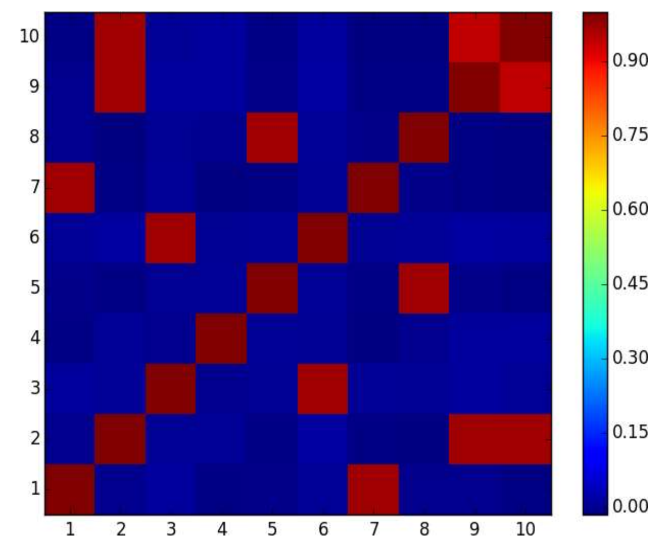
Stage 1: Tree Clustering

- Calculate distance among assets based on their correlation
- Perform hierarchical clustering among assets based on their distances

1. Define the distance measure $d: (X_i, X_j) \subset B \rightarrow \mathbb{R} \in [0,1]$, $d_{ij} = d\,[X_i, X_j] = \sqrt{\frac{1}{2}(1 - \rho_{i,j})}$, where B is the Cartesian product of items in $\{1, \ldots, i, \ldots, N\}$. This forms a proper metric space D
2. Compute the Euclidean distance on D,

$$\tilde{d}: (D_i, D_j) \subset B \rightarrow \mathbb{R} \in [0, \sqrt{N}] = \sqrt{\sum_{n=1}^{N} (d_{n,i} - d_{n,j})^2}$$

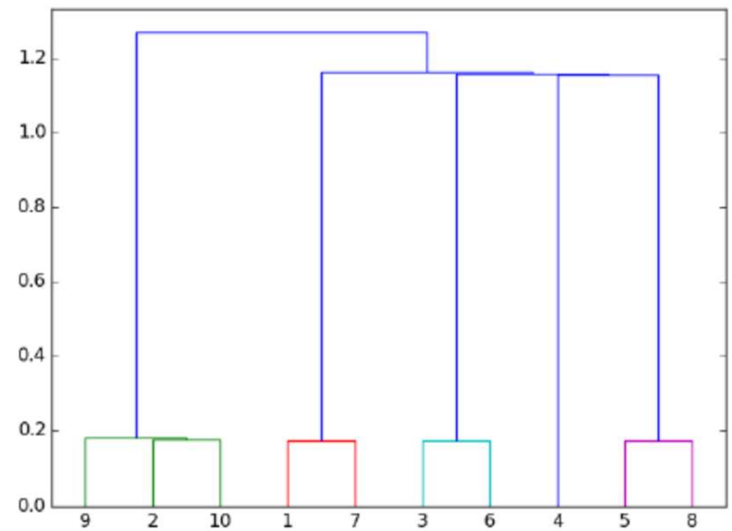
- Note the difference between distance metrics $d_{i,j}$ and $\tilde{d}_{i,j}$. Whereas $d_{i,j}$ is defined on column-vectors of X, $\tilde{d}_{i,j}$ is defined on column-vectors of D (a distance of distances)



Stage 1: Tree Clustering

3. Cluster together the pair of columns (i^*, j^*) such that
$$(i^*, j^*) = \operatorname{argmin}_{\substack{(i,j)\\ i \neq j}} \{\tilde{d}_{i,j}\}$$
4. Update $\{\tilde{d}_{i,j}\}$ with the new cluster
5. Apply steps 3-4 recursively until all $N - 1$ clusters are formed

- Similar items are clustered together, in a tree structure where two leaves are bundled together at each iteration
- The dendogram's y-axis reports the distance between the two joining leaves



Stage 1: Tree Clustering

Cluster together the pair of columns (i^*, j^*) such that

$$(i^*, j^*) = \underset{\substack{(i,j) \\ i \neq j}}{\operatorname{argmin}} \{\tilde{d}_{i,j}\}$$

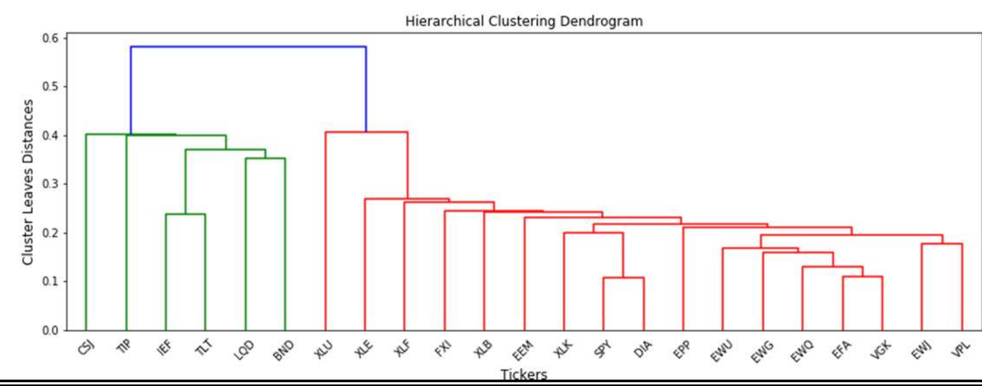
	a	b	c	d	e
a	0	17	21	31	23
b	17	0	30	34	21
c	21	30	0	28	39
d	31	34	28	0	43
e	23	21	39	43	0

	(a,b)	c	d	e
(a,b)	0	**21**	**31**	**21**
c	**21**	0	*28*	*39*
d	**31**	*28*	0	*43*
e	**21**	*39*	*43*	0

Update $\{\tilde{d}_{i,j}\}$ with the new cluster

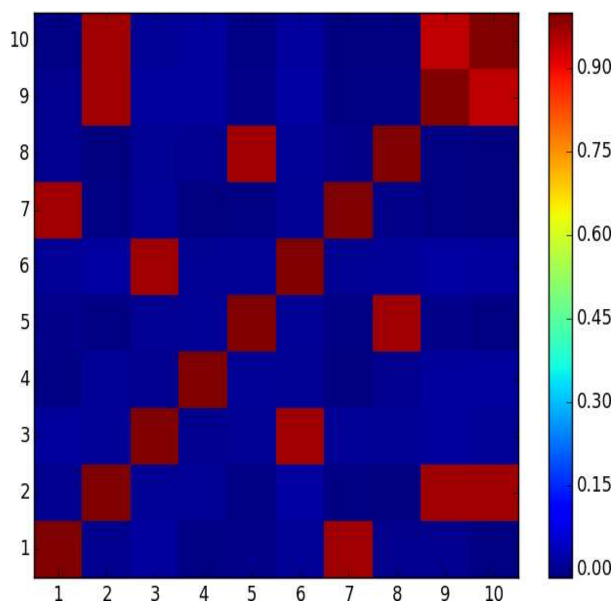
Apply steps 3-4 recursively until all $N-1$ clusters are formed

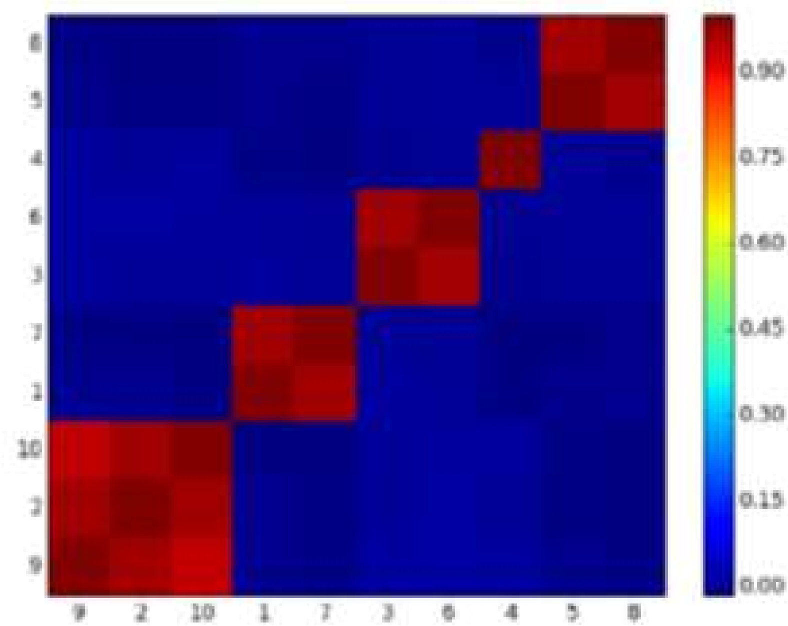
	((a,b),c,e)	d
((a,b),c,e)	0	**28**
d	**28**	0



Stage 2: Quasi-Diagonalization

Sort clustered items by distance using matrix seriation: reorganize the covariance matrix using the results of hierarchical cluster so that similar stocks are together and distant from dissimilar stocks.





GC1

Stage 3: Recursive Bisection

Weights of each cluster of the reordered covariance matrix and of each stock inside each cluster are updated based on the inverse variance.

Steps:

- Recurse from the top to the bottom through the clusters reorganized by the quasi-diagonal matrix where each subcluster of every pair of clusters has the following variance: $V_{adj} = w^T V w$

$$w = \frac{diag[V]^{-1}}{trace(diag[V]^{-1})}$$

- The covariance matrix of each subcluster leads to a new weighting factor that assures a minimum variance portfolio: $\alpha_1 = 1 - \frac{V_1}{V_1 + V_2}; \alpha_2 = 1 - \alpha_1$

- Stocks' weights of left and right subclusters are updated based on the inverse variance:

$$W_1 = \alpha_1 * W_1$$

$$W_2 = \alpha_2 * W_2$$

- As a result, only assets within the same group compete for allocation with each other rather than competing with all the assets in the portfolio.

Slide 15

GC1 German Creamer, 8/8/2020

A Numerical Example

- Simulate a matrix of observations *X*, of order $(10000x10)$
- Add random jumps and a random correlation structure
- Apply three alternative allocation methods:
  - Quadratic optimization, represented by CLA
  - Risk parity, represented by the Inverse Variance Portfolio (IVP)
  - Hierarchical allocation, represented by HRP
- CLA concentrates weights on a few investments, hence becoming exposed to idiosyncratic shocks
- IVP evenly spreads weights through all investments, ignoring the correlation structure:
  -vulnerable to systemic shocks
- HRP diversifies across clusters & items

Weight #	CLA	HRP	IVP
1	14.44%	7.00%	10.36%
2	19.93%	7.59%	10.28%
3	19.73%	10.84%	10.36%
4	19.87%	19.03%	10.25%
5	18.68%	9.72%	10.31%
6	0.00%	10.19%	9.74%
7	5.86%	6.62%	9.80%
8	1.49%	9.10%	9.65%
9	0.00%	7.12%	9.64%
10	0.00%	12.79%	9.61%

Experiment Design

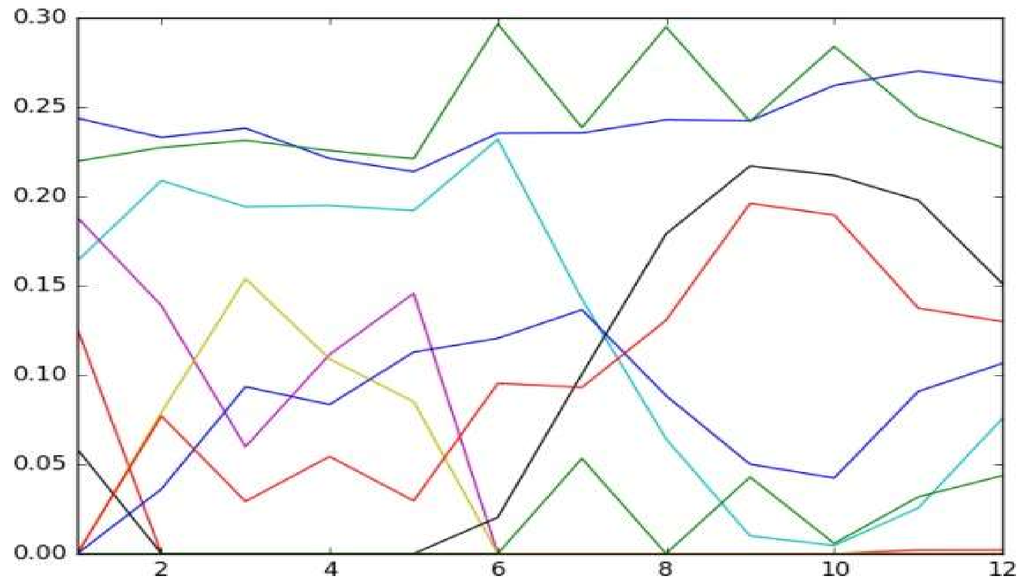
- By definition, CLA has the lowest variance *in-sample*
- **But what method delivers the lowest variance *out-of-sample*?**

1. Generate 10 series of random Gaussian returns (520 observations, equivalent to 2 years of daily history), with 0 mean and an arbitrary standard deviation of 10%
   - Add random shocks, idiosyncratic and common to correlated assets
   - Add a random correlation structure
2. Compute HRP, CLA and IVP portfolios by looking back at 260 observations (a year of daily history)
   - These portfolios are re-estimated and rebalanced every 22 observations (equivalent to a monthly frequency).
3. Compute the out-of-sample returns associated with the three portfolios: CLA, IVP, HRP
4. This procedure is repeated 10,000 times

CLA Allocations

- Variance of the out-of-sample portfolio returns: $\boldsymbol{\sigma^2_{CLA} = .1157}$
- Although **CLA**'s goal is to deliver the lowest variance (that is the objective of its optimization program), its performance happens to exhibit the highest variance *out-of-sample*, and 72.47% greater variance than HRP's

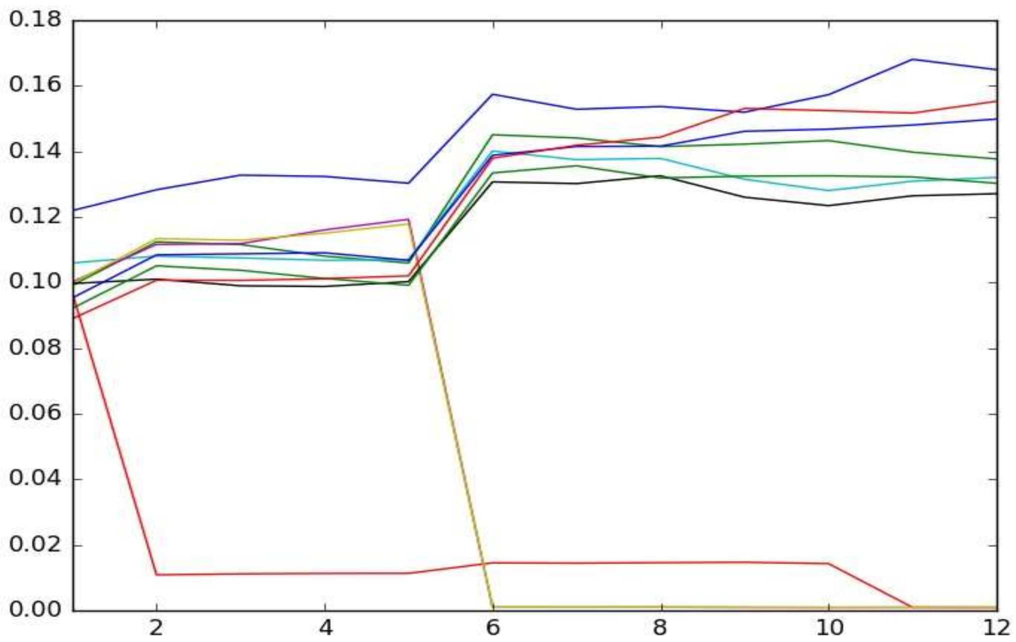
- Let's pick one of the 10,000 experiments, and see how CLA allocations changed between rebalances.
- CLA allocations respond erratically to idiosyncratic and common shocks



IVP Allocations

- Variance of the out-of-sample portfolio returns: $\sigma^2_{IVP} = .0928$
- Assuming that the covariance matrix is diagonal brings some stability to the **IVP**, however its variance is still 38.24% greater than HRP's

- IVP's response to idiosyncratic and common shocks is to reduce the allocation to the affected investment, and spread that former exposure across all other investments
- Consequently, IVP allocations among the unaffected investments grow over time, regardless of their correlation



HRP Allocations

- Variance of the out-of-sample portfolio returns: $\sigma^2_{HRP} = .0671$
- HRP's response to the *idiosyncratic shock* is to reduce the allocation to the affected investment, and use that reduced amount to increase the allocation to a *correlated* investment that was unaffected
- As a response to the *common shock*, HRP reduces allocation to the affected investments, and increases allocation to *uncorrelated* ones (with lower variance)
- **Because Risk Parity funds are leveraged, this variance reduction is critical**