Hierarchical Minimum Variance Portfolios

A Unifying Approach using Schur Complements

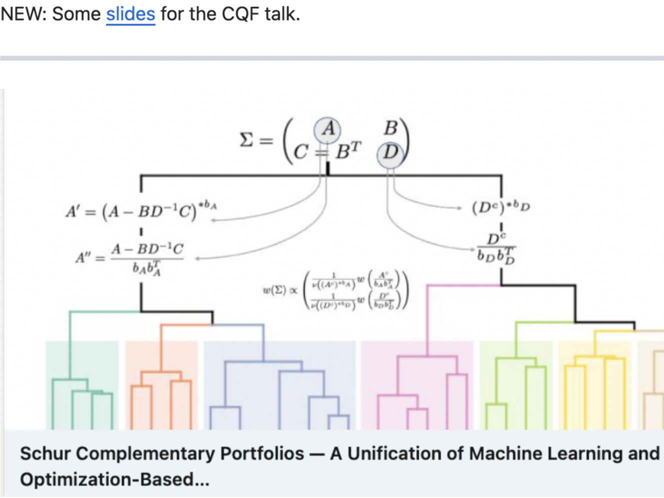
Peter Cotton
Microprediction LLC

Overview

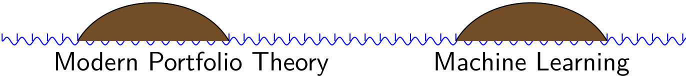
Hierarchical Minimum Variance Portfolios

For these slides, search “microprediction github”.

Then scroll down to the “precise” package ...



This presentation considers two different approaches to allocating capital. Modern Portfolio Theory advocates optimization whereas Lopez de Prado [2] introduced a top-down approach using seriation, and claimed it performs better out of sample.



Which is better?

The answer may surprise you.

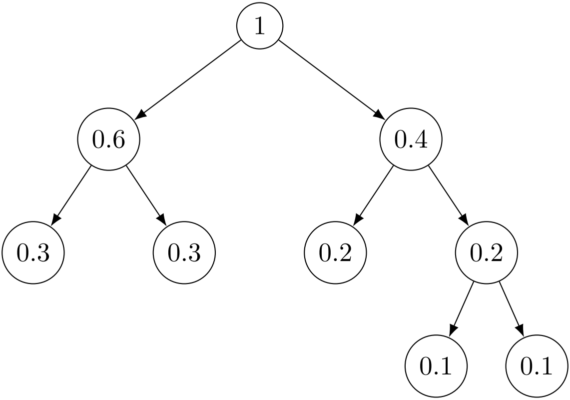
A) The minimum variance portfolio:

$$w \propto \Sigma^{-1} 1 \tag{1}$$

Theoretically the best we can do if Σ is known.

It isn't known.

B) Top-down portfolio construction



If we re-order assets before proceeding to disect the portfolio, we can exploit some information from the covariance matrix but avoid a matrix inversion (Lopez de Prado [2]).

Optimization-based allocation and its discontents

> Of the 14 models we evaluate across seven empirical datasets, none is consistently better than the 1/N rule in terms of Sharpe ratio, certainty-equivalent return, or turnover, which indicates that, out of sample, the gain from optimal diversification is more than offset by estimation error. DeMiguel et al 2009 [3]

> When the objective is to minimize portfolio variance, it turns out mean variance portfolios whose weights are estimated by minimizing the in-sample portfolio variance do not have a superior out-of-sample performance. Prayut and Shashi Jain 2019. [5]

When we compare portfolios associated to multifactor models with mean-variance decisions implied by the single-factor CAPM, we document statistically significant differences in Sharpe ratios of up to 10 percent. Guidolin 2018 [4]

Thus, our results provide an explanation as to why the null hypothesis of equal performance of the simple equally-weighted portfolio compared to many theoretically-superior alternative strategies cannot be rejected in many out-of-sample horse races. Kazak [6]

Monte Carlo experiments show that HRP delivers lower out-of-sample variance than CLA, even though minimum-variance is CLA's optimization objective. De Prado 2016. [2]

Top-down allocation in recursive format.

Algorithm 1 Hierarchical Risk Parity

Input: Covariance matrix $\Sigma = \begin{pmatrix} A & B \\ C = B^T & D \end{pmatrix}$, inter-group allocator ν, intra-group allocator w.

Reorder assets using seriation.

Return $w \propto \begin{pmatrix} \frac{1}{\nu(A)} w(A) \\ \frac{1}{\nu(D)} w(D) \end{pmatrix}$

How can this be justified?

Aside: The financial interpretation of a linear system.

We can view any expression of the form $\Sigma^{-1}\overrightarrow{1}$ in terms of the minimum variance portfolio $w(\Sigma)$ and its portfolio variance $\nu(\Sigma)$, viz:

$$\Sigma^{-1}\overrightarrow{1} = \overrightarrow{1}^T\Sigma^{-1}\overrightarrow{1}\ w(\Sigma) = \frac{1}{\nu(\Sigma)}w(\Sigma)$$

So whenever we see $\Sigma^{-1}\overrightarrow{1}$ appearing in an expression for the overall allocation, a substitution of this style can be made. In particular if $B = 0$ then the global minimum variance allocation is proportional to

$$w \propto \Sigma^{-1}\overrightarrow{1} = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}^{-1}\overrightarrow{1} = \begin{pmatrix} A^{-1}\overrightarrow{1} \\ D^{-1}\overrightarrow{1} \end{pmatrix} = \begin{pmatrix} \frac{1}{\nu(A)}w(A) \\ \frac{1}{\nu(D)}w(D) \end{pmatrix}$$

which is one way to motivate the use of $\nu()$ in a bisection scheme and, one presumes, behind the design in Lopez de Prado [2] not to mention other approaches (e.g. graphs).

But what about $B \neq 0$? Example

$$A = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, B = \begin{pmatrix} \rho \\ \rho \end{pmatrix}, C = \begin{pmatrix} \rho, \rho \end{pmatrix}, D = \begin{pmatrix} 1 \end{pmatrix}$$

Suppose the sub-allocation to assets $\{1, 2\}$ will allocate evenly amongst the two. Then every dollar allocated towards this part of the portfolio, as compared with the third asset, incurs variance

$$v_A = \begin{pmatrix} 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} = \frac{1+\rho}{2}$$

which may alternatively be computed as:

$$1/\nu_A = 1^T A^{-1} 1 = 1^T \frac{\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}}{1-\rho^2} 1 = 2 \frac{(1-\rho)}{1-\rho^2} = \frac{2}{1+\rho}$$

A dollar invested in the third asset incurs unit variance, naturally. Thus a seemingly reasonable top-down portfolio allocation assigning capital inversely proportional to variance will lead to a portfolio allocation

$$\pi_{\{1,2\}} = \frac{\frac{2}{1+\rho}}{\frac{2}{1+\rho}+1} = \frac{2}{3+\rho}$$

and

$$\pi_{\{3\}} = \frac{1}{\frac{2}{1+\rho}+1} = \frac{1+\rho}{3+\rho}$$

Then splitting the allocation to $\{1,2\}$ in half we have

$$w = \frac{1}{3+\rho}\begin{pmatrix} 1 \\ 1 \\ 1+\rho \end{pmatrix}$$

So, despite the reasonableness of this methodology, it evidently over-allocates to asset 3 when $\rho > 0$ and under-allocates when $\rho < 0$.

(I think you saw that coming)

Intermission.

Top-down allocation schemes are potentially pragmatic - not to mention popular. However they are also potentially wasteful of off-diagonal information, and they are definitely unaesthetic.

> Beauty is the first test. There is no permanent place in the world for ugly mathematics.
>
> - G.H. Hardy

Question: Are there beautiful top-down allocation schemes?

Introducing Schur Complementary allocation

A bridge between worlds.

Algorithm 2 Schur Allocation

Input: Covariance matrix $\Sigma = \begin{pmatrix} A & B \\ C = B^T & D \end{pmatrix}$, inter-group allocator ν, intra-group allocator w.

Reorder assets using seriation.

Return $\mathrm{w}(\Sigma; \lambda, \gamma) \propto \begin{pmatrix} \frac{1}{\nu\left((A^c(\gamma))^{*b_A(\lambda)}\right)} w\left(A^c(\gamma)_{/b_A(\lambda)}\right) \\ \frac{1}{\nu\left((D^c(\gamma))^{*b_D(\lambda)}\right)} w\left(D^c(\gamma)_{/b_D(\lambda)}\right) \end{pmatrix}$

---

where terms will be defined.

We don't *simply* use the sub-matrices A and D.

We use augmented matrices which *can*, when we wish, can use information from B.

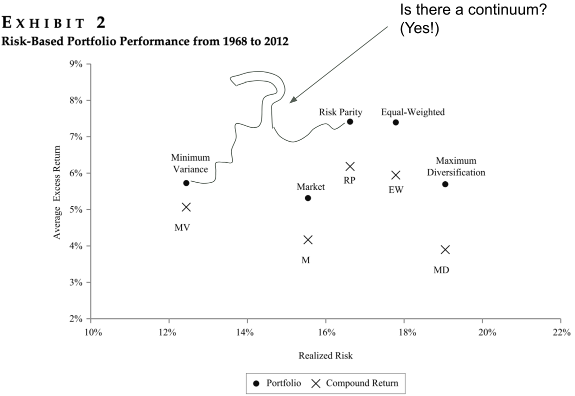


Figure 1: Clarke, de Silva and Thorley 2015 [1]

We will show:

1. Unlike HRP, Schur complementary allocation can reproduce the minimum variance portfolio *exactly*.

You may not *want* to, but it's nice to know you can.

We also observe:

2. Unlike HRP, Schur complementary allocation can use information from B.

Numerical experiments suggest you want that ability.

It also follows:

3. Unlike HRP, Schur complementary allocation can respect *symmetry*.

(How else could it reproduce min var?)

You can trade away your symmetry *consciously*.

Motivating a new family of top-down allocation schemes.

(Similar to the motivation of other methods, if you think about it, but just a bit more careful).

Consider the following standard matrix inversion identity:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{pmatrix} \begin{pmatrix} 1 & -BD^{-1} \\ -CA^{-1} & 1 \end{pmatrix}$$

In our usage $B = C^T$, for now.

We will need the fact that the minimum variance portfolio

$$w^* = \arg\min_w w^T Q w \ s.t. w^T b = 1 \tag{2}$$

is solved by

$$w(Q,b) = \frac{Q^{-1}b}{b^T Q^{-1} b} \tag{3}$$

and the portfolio variance is

$$\nu(Q,b) = w(Q,b)^T Q w(Q,b) = \frac{1}{b^T Q^{-1} b} \tag{4}$$

(Again: we're interested in this because we're trying to first derive a top-down allocation scheme which *can* reproduce the min-var portfolio, quite separately from the question of whether one would wish to invest one's 401k on it without further consideration).

Second aside on linear systems.

Just a slight generalization of the case we saw before when we considered $\Sigma^{-1}1$.

Setting caveats aside we can formerly express the solution to the symmetric linear system $Qx = b$ as:

$$Q^{-1}b = \frac{1}{\nu(Q,b)} w(Q,b) \tag{5}$$

where we use the portfolio optimization solution 3, viz:

$$w(Q,b) = \frac{Q^{-1}b}{b^T Q^{-1} b}$$

So again we can play the game where we spot terms like $Q^{-1}b$ and replace them with a financial interpretation that uses $w(Q,b)$ and the resulting portfolio variance $\nu(Q,b)$.

That is the trick for designing elegant top-down allocation systems.

We're not quite done however because we'll often want to express the solution to this slightly more general portfolio problem in terms of the unconstrained minimum variance portfolio.

That's a change of variables, as I'm sure you anticipated:

$$w(Q,b) = \frac{w(\frac{Q}{bb^T},1)}{b}$$

We can relate the two portfolio variances using

$$b^T Q^{-1} b = 1^T \left(Q^{-1} \cdot (bb^T)\right) 1^T = 1^T \left(\left(Q^{-1} \cdot (bb^T)\right)^{-1} \right)^{-1} 1$$

where the right hand side is the portfolio variance for an augmented covariance matrix. Continuing the algebra...

If we have at our disposal a portfolio construction method $w : Q \rightarrow w(Q)$ generating weights w summing to unity, and some estimate of portfolio variance estimator ν for the same (which might be bravely generalized to other metrics) then

$$Q^{-1}b \leftrightarrow \frac{1}{\nu(Q^{*b})}\frac{w(Q_{/b})}{b} \tag{6}$$

where the notation suggests "might be swapped out for" based on equality in the case of minimum variance portfolios. Here

$$Q_{/b} := \frac{Q}{bb^T} \tag{7}$$

and we also introduced an operator that is a conjugation of matrix inversion with point-wise multiplication by bb^T:

$$Q^{*b} := \left(Q^{-1} \cdot (bb^T)\right)^{-1} \tag{8}$$

We read this operation as "element-wise multiplication in the precision domain".

Now returning to the minimum variance portfolio the matrix inversion identity simplifies to

$$w \propto \Sigma^{-1}\overrightarrow{1} \propto \begin{pmatrix} (A^c)^{-1}\left(\overrightarrow{1} - BD^{-1}\overrightarrow{1}\right) \\ (D^c)^{-1}\left(\overrightarrow{1} - AC^{-1}\overrightarrow{1}\right) \end{pmatrix}$$

So let's denote:

$$b_A(\lambda) := \overrightarrow{1} - \lambda BD^{-1}\overrightarrow{1}$$

and

$$b_D(\lambda) := \overrightarrow{1} - \lambda AC^{-1}\overrightarrow{1}$$

Recall that BD^{-1} appears in the regression of one group against the other, and here comes the financial interpretation (finally):

The inversion identity is trying to stop us from over-investing in pairs of assets from different groups that are strongly correlated (notwithstanding the seriation driving $B \to 0$).

Is it time to tie a bow on this?

We have

$$w \propto \begin{pmatrix} (A^c(\gamma=1))^{-1} b_A(\lambda=1) \\ (D^c(\gamma=1))^{-1} b_D(\lambda=1) \end{pmatrix} \tag{9}$$

but using our "financial interpretation of linear algebra" a top-down allocation procedure can now be "read" (i.e. can be *suggested*) from 9 using 6

$$w(\Sigma;\lambda,\gamma) \propto \begin{pmatrix} \frac{1}{\nu\left((A^c(\gamma))^{*b_A(\lambda)}\right)} w\left(A^c(\gamma)_{/b_A(\lambda)}\right) \\ \frac{1}{\nu\left((D^c(\gamma))^{*b_D(\lambda)}\right)} w\left(D^c(\gamma)_{/b_D(\lambda)}\right) \end{pmatrix} \tag{10}$$

Evidently I've included parameters λ, γ so that 10 need not recreate the minimum variance portfolio.[1]

[1] Not the only way to do this parameterization.

Summarizing Schur top-down allocation.

It *can* be similar to Hierarchical Risk Parity (and use seriation and other ideas) but it is only identical if we set $B = 0$ and we need not do that.

1. The intra-group allocation pertaining to block A is determined by covariance matrix $A^c_{/b_A(\lambda)}$. In this notation the vector $b_A(\lambda) = \overrightarrow{1} - \lambda B D^{-1} \overrightarrow{1}$. The generalized Schur complement is $A^c(\gamma) = A - \gamma B D^{-1} C$. The notation $A^c_{/b}$ denotes $A^c/(bb^T)$ with division performed element-wise.

2. Before performing inter-group allocation we make a different modification. We multiply the *precision* of A^c by $b_A b_A^T$ element-wise (and similarly, multiply the precision of D^c by $b_D b_D^T$).

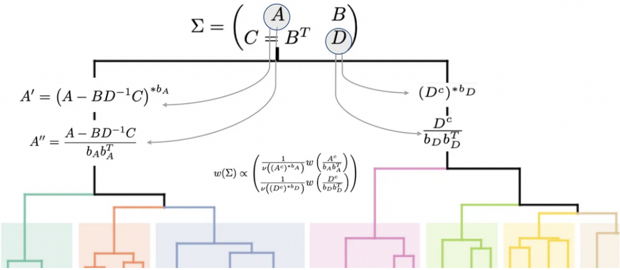
On the other hand it can *also* recreate the min-var portfolio, if the matrix Σ allows the stated computations.

	Top-down	Schur
Inter-group	A or $diag(A)$	$(A^c(\gamma)^{-1} \cdot b_A b_A^T)^{-1}$ $A^c(\gamma) = A - \gamma B D^{-1} C$ and $b_A(\lambda) = \overrightarrow{1} - \lambda B D^{-1} \overrightarrow{1}$.
Intra-group	A	$\left(A - \gamma B D^{-1} C\right) / \left(b_A b_A^T\right)$

Summarizing the distinction between Hierarchical Risk Parity, which merely uses the sub-covariance matrices A and D, versus a family of Schur top-down schemes where A and D are modified as they are passed down through the bisection step.

You can still use structure.

But we have motivated a reason to go beyond the simple use of sub-covariance matrices.



A bit of fun with the M6 Competition

What better way to test the method than an open, year-long, world-wide portfolio contest?

(The M6 Financial Forecasting Competition)

Approximately 200 teams from dozens of countries competed.

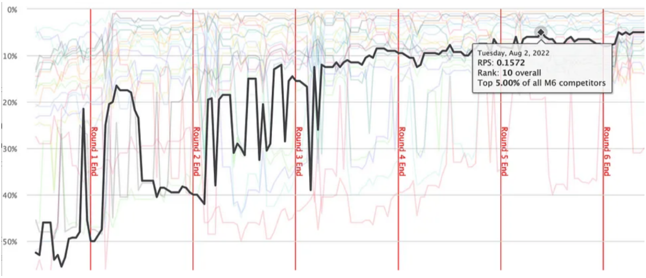
I used Schur portfolio construction applied to a covariance matrix informed by implied volatility.

Aside: Why use implied volatility to guide covariance estimation?

Ans: Markets aren't as dumb as your estimator.

But I'm not here to give investment advice

I merely note that my entry steadily climbed the rankings of the M6 Competition:



Here are the top seven teams very close to the end of the contest. My entry is 7th. How many of those ahead of me have plausibly repeatable Sharpe ratios year on year?

Position	Team	Overall Rank (OR)	Performance of Forecasts (RPS)	Rank (Forecasts)	Performance of Decisions (IR)	Rank (Decisions)
1	cd597d34 StanekF_STU (CERGE-EI)	6	0.15735	6	13.39013	6
2	38c7fc7b MP - Miguel Pérez Michaus	13	0.15661	1	3.65944	25
3	08986844 Peters_STU	16.5	0.15801	11	4.77344	22
4	dd543bd7 Innovation Team (Mathco)	22	0.15922	18	3.00801	26
5	1d3cd369 Rogal Dorn	22.5	0.15968	27	6.58754	18
6	b17cd734 QuantM6.ai	25	0.15979	30	4.97293	20
7	71b47b97 microprediction	27.5	0.15850	13	0.76081	42

None of them.

Does M6 prove a lot in this instance?

Not really. I prefer painfully slow simulated experiments.

Code for Schur portfolios is on GitHub

To be precise ...

https://github.com/microprediction/precise

(where you'll also find my eclectic collection of incremental covariance estimators, related literature reading list, and these slides)

pip install precise

Thanks to Intech Investments and to Adrian Banner, Jian Tang and Jose Marques in particular for feedback.

Marcos Lopez de Prado's paper, cited, motivated me to take a closer look at hierarchical methods in the first place.

A shout to Hugo Delatte, Dany Cajas, Robert Martin, Daniel Palomar and everyone else building open source portfolio packages.

Thanks also to Fred Viole, Dan Pirjol, Peter Schwendner, Jochen Papenbrock, Alejandro Rodriguez Dominguez, Arthur Berd, JP Opdyke, Marco Gorelli, M6 Participants and many others for comments.

(There is some discussion here and there on LinkedIn).

References

[1] Roger Clarke, Harindra de Silva, and Steven Thorley. Risk Parity, Maximum Diversification, and Minimum Variance: An Analytic Perspective. *SSRN Electronic Journal*, 2012.

[2] Marcos López De Prado. Building diversified portfolios that outperform out of sample. *Journal of Portfolio Management*, 42(4), 2016.

[3] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5), 2009.

[4] Massimo Guidolin, Erwin Hansen, and Martín Lozano-Banda. Portfolio performance of linear SDF models: an out-of-sample assessment. *Quantitative Finance*, 18(8), 2018.

[5] Prayut Jain and Shashi Jain. Can machine learning-based portfolios outperform traditional risk-based portfolios? The need to account for covariance misspecification. *Risks*, 7(3), 2019.

[6] Ekaterina Kazak and Winfried Pohlmeier. Testing out-of-sample portfolio performance. *International Journal of Forecasting*, 35(2), 2019.