SCHUR COMPLEMENTARY PORTFOLIOS

Some new approaches to robust asset allocation and covariance estimation are described. In particular, two novel ideas are presented that are easily combined: Schur complementary top-down allocation and "weak" bottom-up portfolio construction. The former uses matrix inversion identities to motivate a family of methods that use divide-and-conquer, yet can mimic global minimum variance optimization. The latter is a simple shrinkage method minimizing the variance of a portfolio whose negative mass has been redistributed.

1. Outline. One goal here is to illustrate convergence between what we term bottom-up allocation, characterized by a single explicit optimization, and top-down portfolio construction (a popular example of the latter is the method of hierarchical risk parity). Throughout, results pertaining to the former will be used to motivate new ideas for the latter.

In recent years superior out of sample performance has been reported for divide-and-conquer capital allocation schemes, as compared with determining all weights with a single optimization whose inputs may be noisy. But this empirical success only highlights the unsatisfying theoretical gap between different philosophies.

In Section 2 we review unit long-short portfolios (weights summing to unity) that minimize portfolio variance given perfect knowledge of covariance. In Section 3 we juxtapose these allocations against those achieved using more heuristic divide-and-conquer approaches, where the latter are viewed as "incorrect" notwithstanding their alleged empirical utility.

In Section 4 these imperfections, so defined, are viewed through the lens of a matrix inversion identity with the goal of convincing the reader that there are sensible covariance sub-matrix augmentations that might enhance top-down methodologies. The message is that top-down portfolio allocation should not merely rely on sub-matrices of the original covariance matrix. The "divide" step in the divide-and-conquering can extend beyond selection of diagonal blocks and in doing so, capture more information.

In particular the use of Schur complements, and modifications of the same, is shown to eliminate the discrepancy between heuristic top-down portfolios and those obtained by minimizing the total portfolio variance. So emboldened, we consider a class of top-down portfolio construction methodologies characterized less by the method that is applied to sub-groups, or the metric used to allocate capital between them, but rather by the nature of the transformations of sub-covariance matrices that get passed down at each recursive step.

Section 5 and Section 6 explore slightly different top-down approaches, each of which can span the seeming divide between this style of allocation and the optimization approach. Parameters are introduced to weaken the use of the Schur complement. As those parameters move from zero to unity, we traverse from traditional top-down allocation such as risk parity to direct optimization.

Naturally the closer we get to optimization, the greater the empirical danger (we are warned). So in Section 13 a comparison is performed using several decades of stock market return data and recommendations are made for a reasonable compromise between the two philosophies.

2. Bottom up allocation. Consider three assets whose returns have equal mean and covariance

$$\Sigma = \Sigma_3 = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

A classic method of allocating capital to those three assets treats it as a constrained minimization of portfolio variance:

$$w(\Sigma) = \arg min_w w^t \Sigma w \quad s.t. \quad w^T \overrightarrow{1} = 1 \tag{2.1}$$

where for avoidance of doubt

$$\overrightarrow{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The solution is well known.

$$w\left(\Sigma, 1\right) = \frac{\Sigma^{-1}\overrightarrow{1}}{\overrightarrow{1}^T \Sigma^{-1} \overrightarrow{1}} \tag{2.2}$$

For brevity we'll write

$$w(\Sigma, 1) = w(\Sigma)$$

to represent the long-short portfolio. This setup is somewhat more general than it may appear, for two reasons. Firstly:

$$w(\Sigma, b) = \arg min_w w^t \Sigma w \quad s.t. \quad w^t b = 1 \tag{2.3}$$

can be reduced to 2.1 by change of variables $v = w \cdot b$. Then the solution is

$$w(\Sigma, b) = \frac{w(\Sigma_{/b}, 1)}{b}$$

where the numerator uses 2.2 and we employ a modified covariance matrix

$$\Sigma_{/b} := \frac{\Sigma}{bb^T}$$

In both cases the fractions represent element-wise division.

Secondly, the equation 2.2 may be employed even if Σ is not a covariance matrix. There is no requirement that it be symmetric, although typically $w^T \Sigma w > 0$.

We use the example Σ_3 as it is obvious, by symmetry considerations alone, that solving 2.1 should result in $w = (1/3, 1/3, 1/3)$. Indeed the application of 2.2 seems like hard work but nonetheless it will be useful to have an explicit representation on hand in what follows. So we observe:

$$\Sigma^{-1} = \frac{1}{1 - \rho - 2\rho^2} \begin{pmatrix} 1+\rho & -\rho & -\rho \\ -\rho & 1+\rho & -\rho \\ -\rho & -\rho & 1+\rho \end{pmatrix} \tag{2.4}$$

and then, continuing to ignore symmetry, diligently compute

$$\nu(\Sigma) := \frac{1}{\overrightarrow{1}^t \Sigma^{-1} \overrightarrow{1}} = \frac{1}{\overrightarrow{1}^T \frac{1-\rho}{1+\rho-2\rho^2} \overrightarrow{1}} = \frac{1}{3} \frac{1+\rho-2\rho^2}{1-\rho}$$

which in generality is the portfolio variance when 2.1 is satisfied. This can be seen from 2.2, by writing $w_1 = w(\Sigma, 1)$ for brevity and noting:

$$w_1^T \Sigma w_1 = w_1^T \frac{\overrightarrow{1}}{\overrightarrow{1}^T \Sigma^{-1} \overrightarrow{1}} = \frac{\overrightarrow{1}^T \Sigma^{-1}}{\overrightarrow{1}^T \Sigma^{-1} \overrightarrow{1}} \frac{\overrightarrow{1}}{\overrightarrow{1}^T \Sigma^{-1} \overrightarrow{1}} = \frac{\overrightarrow{1}}{\overrightarrow{1}^T \Sigma^{-1} \overrightarrow{1}}$$

Finally,

$$w_1 = \nu(\Sigma)\Sigma^{-1}\overrightarrow{1} = \frac{1}{3}\frac{1}{1-\rho}\begin{pmatrix} 1-\rho \\ 1-\rho \end{pmatrix} = \frac{1}{3}\overrightarrow{1}$$

as expected.

In practical settings there may be other constraints other than represented in 2.1 or 2.3 but a class of bottom-up methods all follow a similar routine. We assign characteristics to all constituents in the portfolio and then perform a singular optimization which determines every asset weight at once.

The pitfalls of optimization-based portfolio construction are not intended to spring from these examples, although there may be a hint in the limit $\rho \to 1$. To anticipate the need for alternatives the reader should rather assume Σ is much larger and rank deficient. For instance it might be a $p = 500 \times p = 500$ covariance matrix estimated using only $n = 200$ historical data points. Then, while the formula 2.2 is elegant, its literal use may not be advisable in all cases.

To the contrary, empirical evidence has been presented by numerous authors suggesting that out of sample portfolio performance (as might be measured by inverse variance in our example) can be poor. They have sought approaches that are not predicated on the existence of Σ^{-1}, and that are not subject to instability of solutions arising otherwise for 2.1.

3. Top down portfolio construction. So motivated, we consider a class of top-down portfolio allocation schemes, epitomized by a splitting of capital amongst two mutually exclusive groups of assets and subsequent, independent sub-group allocations.

Typically the inter-group allocation step employs some measure of fitness for receiving capital that is a function of a sub-group covariance matrix.

A brief justification can lean on the solution 2.2. We can view any expression of the form $Q^{-1}\overrightarrow{1}$ in terms of the minimum variance portfolio $w(Q)$ and its portfolio variance $\nu(Q)$, viz:

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

So whenever we see $Q^{-1}\overrightarrow{1}$ 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

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

In words, we allocate inter-group according to the inverse variance of the sub-portfolios in the ratio $1/\nu(A) : 1/\nu(D)$. Then for each group this allocation is further split using the application of some existing portfolio method, or methods, $A \to w(A)$ and $D \mapsto w(D)$. (Although derived using 2.2 the form taken by this expression suggests generalizations beyond the case where the portfolio construction method w is minimum variance.)

Potential weaknesses with this scheme might relate to the assumption $B = C^T = 0$, obviously. But let us walk through the first example with three assets. We might partition the assets into $\{1,2\}, \{3\}$ and, consequently, break down the covariance into

$$\Sigma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where

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

Suppose next that 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:

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

as we have seen.

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 $\rho < 0$.

There are some fixes that apply to this example, it must be said. First, we could average over partitions. Taking an average of the top-down portfolio allocations that use $\{1\}, \{2,3\}$, $\{1,2\}, \{3\}$ and $\{1,3\}, \{2\}$ will restore symmetry as required by this particular example (but not all, unfortunately).

Another approach uses "diagonal" allocation. By this we refer to a calculation that ignores off-diagonal entries in the sub-covariance matrices. This will usually lead to allocations $\pi_{\{1,2\}} = 2/3$ as required - for instance if we do the obvious thing and retain the individual asset variances while discarding covariance.

However, diagonal allocation fails to address the example:

$$\Sigma = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

because it will result in an evenly split allocation $w = \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}^T$ potentially quite different to the optimal solution. More generally diagonal allocation will over-allocate to the sub-portfolios with highest internal correlation, and attract dollars where diversification is in part illusory.

Rather than using bisection, top-down approaches might be suggested by considering each sub-portfolio to be an index or ETF, and computing the inter-group covariance matrix characterizing those aggregated securities. Then, a portfolio construction method could be employed at the level of the aggregates, and that will retain *some* of the information from B and D. We call this the indexing method.

For example if

$$\Sigma = \begin{pmatrix} 1 & \rho & 0 & 0 \\ \rho & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho \\ 0 & 0 & \rho & 1 \end{pmatrix}$$

and we view $\{1,2\}$ and $\{3,4\}$ as single assets with variances $\frac{1+\rho}{2}$, then evidently we allocate evenly and the eventual portfolio is $w = \left(1/4, 1/4, 1/4, 1/4\right)^T$. One might also encounter

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & \rho \\ 0 & 1 & \rho & 0 \\ 0 & \rho & 1 & 0 \\ \rho & 0 & 0 & 1 \end{pmatrix}$$

Again $w = \left(1/4, 1/4, 1/4, 1/4\right)^T$ and if

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho & 0 \\ 0 & 1 & 0 & \rho \\ \rho & 0 & 1 & 0 \\ 0 & \rho & 0 & 1 \end{pmatrix}$$

we are fine. However indexing might run into difficulty in the case

$$\Sigma = \Sigma_4 = \begin{pmatrix} 1 & \rho & \rho & 0 \\ \rho & 1 & \rho & 0 \\ \rho & \rho & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

as we already know that

$$\Sigma^{-1} = \frac{1}{\phi} \begin{pmatrix} 1+\rho & -\rho & -\rho & 0 \\ -\rho & 1+\rho & -\rho & 0 \\ -\rho & -\rho & 1+\rho & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\phi(\rho) := 1 + \rho - 2\rho^2$ is the normalizing term we don't need. By 2.2 the portfolio weights are proportional to:

$$w \propto \Sigma^{-1}\overrightarrow{1} \propto \begin{pmatrix} 1-\rho \\ 1-\rho \\ 1-\rho \\ 1 \end{pmatrix}$$

and in particular, the ratio of investment in asset 4 compared to asset 3 is $1/1-\rho$. In contrast the indexing approach with an even partition $\{1,2\},\{3,4\}$ will treat the last two assets equally, and so this clearly will be inconsistent with a bottom-up optimization. (We will return to Σ_4 in Section 4.2 with a resolution).

Proponents of top-down allocation might object to these examples on various grounds, noting that in practice the elimination of certain terms can effect beneficial shrinkage, or otherwise improve the empirical performance in ways that are not yet well understood. Our intention is not to suggest otherwise, and in fact we shall add to the empirical support for top-down allocation momentarily.

(However in the same breath it would be cleanest to separate out approximation and parsimony. And in attempting to ameliorate these theoretically discomforting examples, we might stumble on generalizations of existing top-down approaches that also perform well empirically, or even better.)

We also hasten to add that existing technique can sometimes address these issues. For instance a mere reordering (e.g. using *seriation*) might suffice, and other more careful schemes for clustering assets might reduce or eliminate the mis-allocation. In this last example, the partition $\{1,2,3\},\{4\}$ will solve the problem, for example, and eliminate any discrepancy between the weights deemed optimal with perfect knowledge of Σ and those arrived at by top-down allocation.

A more trivial remark is that some top-down allocation schemes will call down to an optimization routine once the number of assets is reasonably small–potentially making these low-dimensional examples irrelevant in a literal sense. Surely, however, it is clear that that there are block equivalents of these examples, and also that the same ideas are relevant in higher dimensions regardless.

So caveats aside we adopt the applied mathematical standpoint that top-down approaches are crude if empirically effective homogenization. They present a strong motivation to try to unify top-down and bottom-up portfolio construction in a manner that is more satisfying and, hopefully, even more empirically impressive. Referencing this last example, and in contrast to the indexing approach, our ambition might be phrased as a desire to perform top-down allocation in a manner that does not trivially collapse the information in

$$B = \begin{pmatrix} -\rho & 0 \\ -\rho & 0 \end{pmatrix}$$

say. Rather, we will use B in totality in the allocation step.

4. Inversion identity. A family of new top-down approaches will be based on the following standard matrix inversion identity.

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A-BD^{-1}C)^{-1} & 0 \\ 0 & (D-CA^{-1}B)^{-1} \end{pmatrix} \begin{pmatrix} 1 & -BD^{-1} \\ -CA^{-1} & 1 \end{pmatrix} \tag{4.1}$$

(In our usage $B = C^T$ but carrying through their separate identities may save us from error.) We are determined that the troublesome examples mentioned above will not defeat our new method unless we explicitly permit the same. In particular we'll look to ensure that when applied to the examples:

$$\Sigma_4 = \begin{pmatrix} 1 & \rho & \rho & 0 \\ \rho & 1 & \rho & 0 \\ \rho & \rho & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

and the difficult partitions of assets $\{1,2\},\{3,4\}$ and $\{1,2\},\{3\}$ respectively, we can match the global minimum variance solution–at least for some methodological meta-parameters.

To get a sense of what is gained or lost in existing methods, and the new ones suggested, it is helpful to view some special cases of 4.1.

4.1. Three asset example. For Σ_3 we denote

$$\begin{aligned} A^c = A - BD^{-1}C &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \rho \\ \rho \end{pmatrix} (1) \begin{pmatrix} \rho & \rho \end{pmatrix} \\ &= \begin{pmatrix} 1-\rho^2 & \rho-\rho^2 \\ \rho-\rho^2 & 1-\rho^2 \end{pmatrix} = (1-\rho^2) \begin{pmatrix} 1 & \frac{\rho}{1+\rho} \\ \frac{\rho}{1+\rho} & 1 \end{pmatrix} \end{aligned}$$

and note that the inverse is

$$\begin{aligned} (A^c)^{-1} &= \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \frac{\rho}{1+\rho} \\ \frac{\rho}{1+\rho} & 1 \end{pmatrix}^{-1} = \frac{1}{1-\rho^2} \frac{1+\rho}{2\rho+1} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix} \\ &= \frac{1}{1+\rho-2\rho^2} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix} = \frac{1}{\phi(\rho)} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix} \end{aligned}$$

We also have inverse sub-portfolio variance given by

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

A unit sub-portfolio determined by covariance A^c has weights
(4.2)

$$w(A^c,1) = \frac{(A^c)^{-1}\overrightarrow{1}}{\overrightarrow{1}^T (A^c)^{-1} \overrightarrow{1}} = \frac{\phi(\rho)}{2} \frac{1}{\phi} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and we'll be using that expression shortly. Similarly

$$D^c := D - CA^{-1}B = 1 - \begin{pmatrix} \rho & \rho \end{pmatrix} \frac{\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}}{1-\rho^2} \begin{pmatrix} \rho \\ \rho \end{pmatrix} = \left(1 - \frac{2\rho^2(1-\rho)}{1-\rho^2}\right) = \left(\frac{\phi(\rho)}{1+\rho}\right)$$

with inverse

$$(D^c)^{-1} = \left(\frac{1+\rho}{\phi(\rho)}\right)$$

of course. (In passing, we can check the portfolio variance

$$\nu_{D^c} = \overrightarrow{1}^T (D^c)^{-1} \overrightarrow{1} = \frac{1+\rho}{\phi(\rho)}$$

which might have been anticipated, given that there is only one possible sub-portfolio). Next the inversion identity is verified. The inverse we expect was given in 2.4, namely:

$$\Sigma^{-1} = \frac{1}{\phi(\rho)} \begin{pmatrix} 1+\rho & -\rho & -\rho \\ -\rho & 1+\rho & -\rho \\ -\rho & -\rho & 1+\rho \end{pmatrix}$$

and evidently $(\Sigma_3^{-1})_{3,3}$ corresponds to $1/D^c$ as required by 4.1. Also

$$-CA^{-1} = -\left(\rho, \rho\right) \frac{\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}}{1-\rho^2} = \left(\tfrac{\rho}{1+\rho}, \tfrac{\rho}{1+\rho}\right)$$

Then

$$-(D^c)^{-1}CA^{-1} = -\frac{1+\rho}{\phi(\rho)} \left(\tfrac{\rho}{1+\rho}, \tfrac{\rho}{1+\rho}\right) = \begin{pmatrix} \frac{-\rho}{\phi(\rho)} & \frac{-\rho}{\phi(\rho)} \end{pmatrix}$$

in accordance with 2.4. Similarly we note

$$-BD^{-1} = -\begin{pmatrix} \rho \\ \rho \end{pmatrix}$$

and

$$-(A^c)^{-1}BD^{-1} = -\frac{1}{1+\rho-2\rho^2} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix}$$

also in accordance with 2.4. Putting these blocks together, the identify 2.4 manifests in the case Σ_3 as

$$\Sigma_3^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A^c)^{-1} & 0 \\ 0 & (D^c)^{-1} \end{pmatrix} \begin{pmatrix} 1 & -BD^{-1} \\ -CA^{-1} & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}^{-1} = \frac{1+\rho}{\phi} \begin{pmatrix} 1 & -\frac{\rho}{1+\rho} & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\rho \\ 0 & 1 & -\rho \\ -\frac{\rho}{1+\rho} & -\frac{\rho}{1+\rho} & 1 \end{pmatrix}$$

We check that weights of the minimum variance portfolio 2.2 are equal:

$$w \propto \Sigma^{-1}\overrightarrow{1} \propto \begin{pmatrix} 1 & -\frac{\rho}{1+\rho} & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\rho \\ 0 & 1 & -\rho \\ -\frac{\rho}{1+\rho} & -\frac{\rho}{1+\rho} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1-\rho}{1+\rho} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Next we shall attempt to layer an interpretation on the matrix identity. Multiplying out from the right and recognizing the role of A^c using 4.2 we can suggestively write:

$$\begin{aligned} w \propto \Sigma^{-1}\overrightarrow{1} &\propto \begin{pmatrix} 1 & -\frac{\rho}{1+\rho} & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1-\rho \\ 1-\rho \\ \frac{1-\rho}{1+\rho} \end{pmatrix} \\ &\propto \begin{pmatrix} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 2\ w(A^c) \\ w(D^c) \end{pmatrix} \end{aligned} \tag{4.3}$$

Although this is the bottom-up solution, and of course trivial by symmetry, the intermediate steps in the calculation we are highlighting can be read in top-down fashion. We allocate to assets $\{1,2\}$ and $\{3\}$ in the ratio $2:1$ and then split between 1 and 2 using the minimum variance portfolio determined by the Schur complement A^c. Due to cancellation of terms, the allocation between $\{1,2\}$ and $\{3\}$ is correct now, unlike when $1/\nu(A) : 1/\nu(D)$ is employed in a more traditional top-down approach.

4.2. Four asset example. In a similar vein we can approach the example

$$w \propto \Sigma^{-1}\overrightarrow{1} = \Sigma = \begin{pmatrix} 1 & \rho & \rho & 0 \\ \rho & 1 & \rho & 0 \\ \rho & \rho & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \overrightarrow{1}$$

with the equal but somewhat troublesome partition

$$A = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \; B = \begin{pmatrix} \rho & 0 \\ \rho & 0 \end{pmatrix}, C = \begin{pmatrix} \rho & \rho \\ 0 & 0 \end{pmatrix}, \; D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This time

$$A^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

$$-CA^{-1} = -\frac{\rho}{1+\rho} \begin{pmatrix} \rho & \rho \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} = -\frac{2\rho^2(1-\rho)}{1+\rho} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

Also

$$-BD^{-1} = -B = \begin{pmatrix} -\rho & 0 \\ -\rho & 0 \end{pmatrix}, \; D^{-1} = D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

And Schur complements are

$$A^c = A - BD^{-1}C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \rho^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = (1-\rho) \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix}$$

which we've seen before. Again

$$(A^c)^{-1} = \frac{1}{\phi(\rho)} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix}$$

and

$$\begin{aligned} D^c = D - CA^{-1}B &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2\rho^2(1-\rho)}{1+\rho} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \rho & 0 \\ \rho & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{4\rho^3(1-\rho)}{1+\rho} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \frac{4\rho^3(1-\rho)}{1+\rho} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1+\rho-4\rho^3(1-\rho)}{1+\rho} & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

So

$$(D^c)^{-1} = \begin{pmatrix} \frac{1+\rho}{1+\rho-4\rho^3+4\rho^4)} & 0 \\ 0 & 1 \end{pmatrix}$$

Using the matrix inversion identity we have

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

$$= \begin{pmatrix} (A^c)^{-1} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2\rho^2(1-\rho)}{1+\rho} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ (D^C)^{-1} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \rho & 0 \\ \rho & 0 \end{pmatrix} \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{pmatrix}$$

There are two ways these expressions might be made more readable from the perspective of the designer of a top-down allocation strategy. First, we will collapse the square matrices and leave the vector of ones untouched, leading to a method described in Section 5. The alternative route multiplies out the one vectors and will be discussed in Section 6.

Taking the first fork in the road, let us assume there are invertible matrices W_A and W_D satisfying

$$W_A^{-1} = 1 - BD^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2\rho^2(1-\rho)}{1+\rho} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

and

$$W_D^{-1} = 1 - CA^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \rho & 0 \\ \rho & 0 \end{pmatrix} = \begin{pmatrix} 1-\rho & 0 \\ -\rho & 1 \end{pmatrix}$$

Indeed in this example we can write down:

$$W_A = \begin{pmatrix} 0 & 1 \\ -\frac{1}{\rho} & \frac{1-\rho}{\rho} \end{pmatrix}$$

and

$$W_D = \begin{pmatrix} \frac{1+\rho}{1+\rho-2\rho^2+2\rho^3} & -\frac{2(1-\rho)\rho^2}{1+\rho-2\rho^2+2\rho^3} \\ 0 & 1 \end{pmatrix}$$

Then

$$\Sigma^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \propto \begin{pmatrix} (W_A A^c)^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ (W_D D^c)^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{pmatrix} \tag{4.4}$$

has reduced the global allocation to a top-down friendly format.

5. Schur complementary portfolios, method I. Equation 4.4 begs us to write $\tilde{A} = W_A A^c$ and $\tilde{D} = W_D D^c$ where $W_D^{-1} = 1 - CA^{-1}$ and $W_A^{-1} = 1 - BD^{-1}$. For then the minimum variance portfolio weights are proportional to

$$\Sigma^{-1}\overrightarrow{1} \propto \begin{pmatrix} \frac{1}{\nu(\tilde{A})} w(\tilde{A}) \\ \frac{1}{\nu(\tilde{D})} w(\tilde{D}) \end{pmatrix} \tag{5.1}$$

and this would seem to motivate a top-down portfolio approach that uses $\tilde{A}$ and $\tilde{D}$ in place of the customary A and D - even if we do not specifically seek minimum portfolio variance as the only portfolio objective or believe it is likely to be robust.

Actually, we advocate the generalization that replaces $1-BD^{-1}$ with $1-\lambda BD^{-1}$ for some parameter $\lambda \in (0,1)$ and likewise $1-CA^{-1} \to 1-\lambda CA^{-1}$, in order to create a continuum of methods, some relying more heavily on off-block-diagonal terms than others. In similar spirit we introduce a second parameter γ into the construction of the Schur complements

$$A^c(\gamma) := A - \gamma BD^{-1}C$$

There would seem to be some elegance in determining a compelling fixed relationship between γ and λ, but we leave that to empirical study. To summarize:

1. Matrices A and D used in traditional top-down allocation are replaced by their respective augmented Schur complements $\tilde{A} = W_A A^c$ and $\tilde{D} = W_D D^c$ with W_A and W_D chosen to satisfy $W_D(1-CA^{-1}) = 1$ and similarly $W_A(1-BD^{-1}) = 1$.
2. More generally, these auxiliary matrices are determined so that $W_D(1-\lambda CA^{-1}) = 1$ and $W_A(1-\lambda BD^{-1}) = 1$ respectively, introducing a parameter $\lambda \in (0,1)$.
3. More generally, the Schur complement A^c can be replaced by $A - \gamma BD^{-1}C$ introducing a parameter $\gamma \in (0,1)$.

We remark that in the first example Σ_3 the augmented Schur matrix $\tilde{A}^c$ was a multiple of the Schur complement A^c and thus symmetric, but in the second example involving Σ_4 we saw a twisting occur that is necessary to recover the global minimum variance portfolio.

The fact that $\tilde{A}$ need not be symmetric may limit the set of available choices for w and ν, since many portfolio packages assume or enforce symmetry.[1]

6. Schur complementary portfolios, method II. As an alternative to the use of W_A and W_D matrices, we can revisit the relationship between $Q^{-1}b$ and portfolio weights in a minimum variance problem. In particular a solution to

$$w* = \arg\min_w w^T Q w \;\; s.t. w^T b = 1 \tag{6.1}$$

is

$$w(Q,b) = \frac{Q^{-1}b}{b^T Q^{-1} b} \tag{6.2}$$

and the portfolio variance is

$$\nu(Q,b) = w(Q,b)^T Q w(Q,b) = \frac{1}{b^T Q^{-1} b} \tag{6.3}$$

Thus the solution to any symmetric linear system $Qx = b$ (not just the special case $Qx = \overrightarrow{1}$) has a financial interpretation:

$$Q^{-1}b = \frac{1}{\nu(Q,b)} w(Q,b) \tag{6.4}$$

in terms of the portfolio optimization 6.2 yielding $w(Q,b)$ and the resulting portfolio variance $\nu(Q,b)$. We had previous observed:

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

[1]As another minor annoyance, when the number of assets is odd the bisection cannot be perfect and minor adjustment must be made to match dimensions–see the code for details.

and we can also write the portfolio variance in terms of the portfolio variance for the case $b = \overrightarrow{1}$ by noting

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

Therefore if we have at our disposal a portfolio construction method $Q \to w(Q)$ 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

$$Q^{-1} b \leftrightarrow \frac{1}{\nu(Q^{*b})} \frac{w(Q_{/b})}{b} \tag{6.5}$$

where the notation suggests "can be swapped out for" based on equality in the case of minimum variance portfolios. Here we have reused the shorthand

$$Q_{/b} := \frac{Q}{bb^T} \tag{6.6}$$

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

$$Q^{*b} := \left(Q^{-1} \cdot (bb^T)\right)^{-1} \tag{6.7}$$

We read this operation as "element-wise multiplication in the precision domain". Now returning to the minimum variance portfolio we have

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

So denoting

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

and

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

we have

$$w \propto \begin{pmatrix} (A^c(\gamma = 1))^{-1} b_A(\lambda = 1) \\ (D^c(\gamma = 1))^{-1} b_D(\lambda = 1) \end{pmatrix} \tag{6.8}$$

and thus the following top-down portfolio family can be "read" using 6.5

$$w(\Sigma; \lambda, \gamma) \propto \begin{pmatrix} \frac{1}{\nu\left((A^c(\gamma))^{*b_A(\lambda)}\right)} w\left(A^c(\gamma)_{/b_A(\lambda)}\right) \\ \frac{1}{\nu\left((D^c(\gamma))^{*b_D(\lambda)}\right)} w\left(D^c(\gamma)_{/b_D(\lambda)}\right) \end{pmatrix} \tag{6.9}$$

Once again this format is highly suggestive. We can be empirically guided in our choice of w and ν (reaching for whatever we might find in portfolio allocation packages) even though the derivation was only for minimum variance portfolios.

If $w(Q, 1)$ is any method of determining a portfolio with $w^t \overrightarrow{1} = 1$ and $\nu(Q)$ is some assessment (similar to portfolio variance in spirit but not necessarily the same) then the same steps can be applied. Some possibilities are listed in Table 2

Allocation	Risk-parity	Schur II
Inter-group	A or $diag(A)$	$(A^c(\gamma)^{-1} \cdot b_A b_A^T)^{-1}$ where we set $A^c(\gamma) = A - \gamma BD^{-1}C$ and $b_A(\lambda) = \overrightarrow{1} - \lambda BD^{-1}\overrightarrow{1}$.
Intra-group	A	$\left(A - \gamma BD^{-1}C\right) / \left(b_A b_A^T\right)$ element-wise division

TABLE 1

Comparison of covariance matrices used to allocate capital between and within a group of securities in two different divide-and-conquer schemes.

For instance $1/\nu(Q)$ could be any allocation scheme provided the ratio $1/\nu(A)$: $1/\nu(D)$ would ordinarily be considered a sensible splitting of funds. We have used the case of the minimum variance portfolio to motivate the use of a different splitting $1/\nu((A^c)_{/b_A}) : 1/\nu((D^c)_{/b_D})$ that could employ the same metric but different matrices as input.

As with Section 5 we emphasize again that this approach differs from existing top-down allocation methods only in the "divide" step, and specifically in the matrices passed from one function to the next. Summarizing:

1. The intra-group allocation pertaining to block A is determined by covariance matrix $A^c_{/b_A(\lambda)}$. In this notation the vector $b_A(\lambda) = \overrightarrow{1} - \lambda BD^{-1}\overrightarrow{1}$. The generalized Schur complement is $A^c(\gamma) = A - \gamma BD^{-1}C$. 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 $b_D b_D^T$).

Unlike Section 5 the modified versions of A and D are symmetric. See also Table 1 for a comparison against risk-parity.

In some cases the choice of γ might be suggested by numerical properties of the matrices created–such as by choosing $\gamma < 1$ to ensure that A^c is positive definite.

The continuum established is hopefully clear from Table 3 and the definitions $A^c(\lambda) = A - \lambda BD^{-1}C$ and $D^c(\lambda) = D - \lambda CA^{-1}B$. Also $b_A(\gamma) := \overrightarrow{1} - \gamma BD^{-1}\overrightarrow{1}$ and $b_D(\gamma) := \overrightarrow{1} - \gamma AC^{-1}\overrightarrow{1}$. Clearly for $\gamma \to 0$ and $\lambda \to 0$ we recover the usual A and D using top-down allocation, thereby recovering oft-used allocation schemes that ignore $B = C^T$, such as hierarchical risk parity.

Whereas our approach has emphasized that as $\gamma \to 1$ and $\lambda \to 1$ we can, modulo linear algebra difficulties, hope to fully reproduce the globally minimum variance portfolio. Thus γ and λ can parameterize the space between top-down and bottom-up portfolios, or at least some part of it reaching to minimum variance unit portfolios. No suggestion is made that Schur top-down allocation can alleviate numerical or empirical pains in the region close to $\gamma = \lambda = 1$, however.

7. Schur complementary portfolios III: leave one out. The matrix identity in Section 4 is not the only inspiration for modified top-down approaches. In this Section we illustrate a different identity and a possible use. This time we consider the following special case of an $n+1 \times n+1$ covariance matrix:

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma & \eta\beta \cdot \sigma \\ \eta\beta \cdot \sigma & \eta^2 \end{pmatrix}$$

for some scalar η and n-vectors β and σ describing assets and their relationship to the $n+1$'st asset. Here the sub-covariance matrix Σ's entries are given by

$$\Sigma_{i,j} = \rho_{i,j}\sigma_i, \sigma_j$$

In one interpretation the $n+1$'st asset is an index made up of combinations of the other n assets, so that $\tilde{\Sigma}$ is rank deficient.

As with previous sections we shall reformulate the unit minimum variance portfolio:

$$\tilde{w} \propto \tilde{\Sigma}^{-1}\overrightarrow{1}$$

This time we employ an inversion identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -D^{-1}C & 1 \end{pmatrix} \begin{pmatrix} (A-BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} 1 & -BD^{-1} \\ 0 & 1 \end{pmatrix} \tag{7.1}$$

with $A = \Sigma$, $D = \eta^2$, $B = \eta\beta \cdot \sigma = C^T$. Used as:

$$\tilde{\Sigma}^{-1}\overrightarrow{1} = \begin{pmatrix} 1 & 0 \\ -\beta\cdot\sigma/\eta & 1 \end{pmatrix} \begin{pmatrix} (\Sigma^c)^{-1} & 0 \\ 0 & \eta^{-2} \end{pmatrix} \begin{pmatrix} 1 & -\beta\cdot\sigma/\eta \\ 0 & 1 \end{pmatrix} \overrightarrow{1}$$

where

$$\Sigma^c_{i,j} = \sigma_i\sigma_j\left(\rho_{i,j} - \beta_i\beta_j\right) \tag{7.2}$$

represents covariance not explained by "beta". Then if we denote

$$b = \overrightarrow{1} - \beta\cdot\sigma/\eta \tag{7.3}$$

we have

$$\tilde{\Sigma}^{-1}\overrightarrow{1} = \begin{pmatrix} 1 & 0 \\ -\beta\cdot\sigma/\eta & 1 \end{pmatrix} \begin{pmatrix} (\Sigma^c)^{-1}b \\ 1/\eta^2 \end{pmatrix}$$

So as with 6.5 if

$$\Sigma^{*b} := \left((\Sigma^c)^{-1}bb^T\right)^{-1}$$

Then we see the allocation can be split between the n assets and the $n+1$'st, since

$$w \propto \begin{pmatrix} 1 & 0 \\ -\beta\cdot\sigma/\eta & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\nu(\Sigma^{*b})}\frac{w\left(\Sigma_{/b}\right)}{b} \\ 1/\eta^2\cdot 1 \end{pmatrix}$$

Multiplying by $\nu(\Sigma^{*b})$ and summing over the first n assets the ratio of investment in the first n assets versus the last is:

$$\begin{pmatrix} \sum_{j=1}^n w_j \\ w_{n+1} \end{pmatrix} \propto \begin{pmatrix} \sum_{j=1}^n \frac{1}{1-\beta_j\sigma_j/\eta} w_j \\ \nu(\Sigma^{*b})/\eta^2 - \sum_{j=1}^n \frac{\beta_j\sigma_j/\eta}{1-\beta_j\sigma_j/\eta} w_j \end{pmatrix} \tag{7.4}$$

where $w_j = w\left(\Sigma_{/b}\right)_j$ and, in full:

$$\Sigma^{*b} = \left(\left(\Sigma - (\overrightarrow{1} - \beta\cdot\sigma/\eta)(\overrightarrow{1} - \beta\cdot\sigma/\eta)^T\right)^{-1} \cdot (\overrightarrow{1} - \beta\cdot\sigma/\eta)(\overrightarrow{1} - \beta\cdot\sigma/\eta)^T\right)^{-1}$$

whereas

$$\Sigma_{/b} := \frac{\Sigma}{bb^T} = \frac{\Sigma}{(\overrightarrow{1} - \beta\cdot\sigma/\eta)(\overrightarrow{1} - \beta\cdot\sigma/\eta)^T}$$

8. Index beating portfolios. If we furthermore assume that the last asset is

$$X_{n+1} = \sum_k \mu_k X_k$$

with $\mu^T \overrightarrow{1} = 1$.

An interesting question that arises is the optimum portfolio in this rank-deficient setting, and whether different ways of approaching the problem will result in the same answer. (Of course the result should be invariant to the addition of subtraction of any multiple of the weight vector

$$\mu^+ = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \\ -1 \end{pmatrix}$$

but setting that symmetry aside, there is no guarantee that any particular approach for dealing with rank-deficiency will seem consistent.)

For brevity write $\nu = \nu(\Sigma^{*b})$. If X_k represent sectors (say) then

$$\beta = (\mu \cdot \sigma/\eta)^T C$$

where C is the sector correlation matrix and

$$\eta^2 = (\mu \cdot \sigma/\eta)^T C \, (\mu \cdot \sigma/\eta) \tag{8.1}$$

So

$$\beta = \frac{(\mu \cdot \sigma/\eta)^T C}{\sqrt{(\mu \cdot \sigma/\eta)^T C \, (\mu \cdot \sigma/\eta)}} \tag{8.2}$$

If μ are known, and $\sigma_1, \ldots, \sigma_n$ and η are also assumed known from option markets, then 8.1 becomes a constraint assisting the estimation of $C_{i,j} = \rho_{i,j}$ and with it the unexplained correlation terms in 7.2, namely

$$\Sigma^c_{i,j} = \sigma_i \sigma_j \left(\rho_{i,j} - \beta_i \beta_j \right)$$

We might write

$$\Sigma^c = \sigma^T U \sigma$$

where U carries the meaning "unexplained" or idiosyncratic. So after the allocation decision is made in 7.4 the covariance used to allocate within the first n assets is

$$\Sigma^c_{/b} = \frac{\sigma^T U \sigma}{(1 - \beta \cdot \sigma/\eta)(1 - \beta \cdot \sigma/\eta)^T} \tag{8.3}$$

Needless to say that if all parameters are known, including U, with precision then this route is circuitous. One reason to proceed this way is that the overall character, or idea, of the long/short portfolio can be preserved but at the same time, the robustifying benefits of long-only portfolio construction can apply assuming only that 8.3 is well behaved. In reading the allocation 7.4 we note

$$\nu(\Sigma^{*b}) < \eta^2$$

Method	$w(Q)$	$\nu(Q)$
Equal	Equal weight portfolio	Portfolio variance
Diagonal	Ignore off-diagonal covariance entries and find minimum variance portfolio	Portfolio variance
Weak	Choose a constant multiplicative reduction in off-diagonal covariance entries to minimize portfolio variance of a long-only renormalized portfolio.	Portfolio variance
Sharpe	Maximum Sharpe ratio assuming fixed mean returns	Portfolio variance

TABLE 2

Some methods used for long-only allocation $w(Q, 1)$ and portfolio assessment $\nu(Q)$

Method	Methodology	Schur equivalent
Hierarchical risk parity	Recursive allocation, typically with “diagonal” allocation.	$\gamma = \lambda = 0$
Minimum variance	One-off global constrained minimization of portfolio variance.	$\gamma = \lambda = 1$

TABLE 3

Extremal points on a continuum of portfolio allocation methods spanning both top-down and bottom-up (optimization) philosophies. Corresponding parameters for Schur complementary portfolios are shown that can, modulo technicalities, replicate the output. This assumes seriation has been used to reorder assets, an important ingredient in hierarchical risk parity.

by definition since the left hand side is a minimum variance portfolio and $\eta^2 = \mu^T \Sigma^{*b} \mu$ is just an example of one portfolio from which this minimum is chosen. So the first term in

$$\nu(\Sigma^{*b})/\eta^2 - \sum_{j=1}^{n} \frac{\beta_j \sigma_j/\eta}{1 - \beta_j \sigma_j/\eta} w_j$$

is less than unity. The second term is typically greater than 1 and certain is if all $\beta_j > 0$. The allocation to the n'th asset is probably negative.

9. Microprediction. However that isn't the case. Instead 8.3 suggests a different path:

1. Determine an idiosyncratic portfolio using 8.3, and possible recursion.
2. Use 7.4 to suggest the overall allocation between constituents and the n'th asset, but...
3. Use microprediction streams to improve on 7.4 by soliciting distributional predictions of several different index allocations.
4. Use microprediction streams to solicit forecasts of β.
5. Independently microprediction streams to solicit forecasts of σ/σ_{imp} where σ_{imp} is the option-implied standard deviation.

The discipline imposed by the M6 Competition provides an opportunity to test these ways of creating index-beating portfolios out of the SPDR sector indexes, and test it only once.

10. Weak portfolios. Here I'm adding some quick thoughts about weak portfolios, as they are called in the code. The empirical performance has been surprisingly good despite the simplicity of the approach, which is an adaptive ridge. Here's an

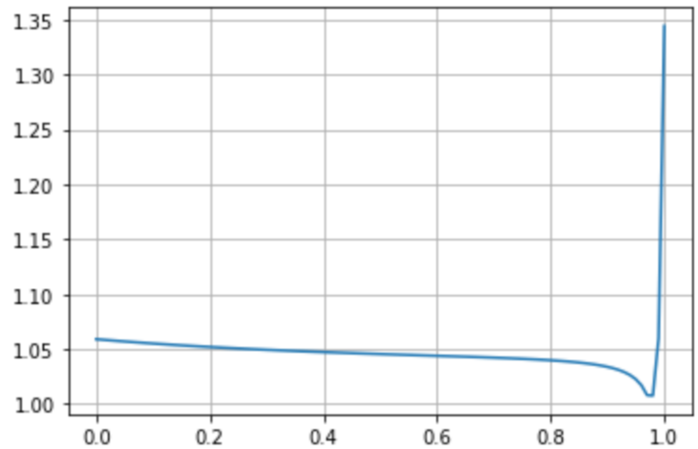

FIG. 1. *Portfolio variance of the long-only portfolio that is achieved by first multiplying off-diagonal entries and then redistributing the mass. The minimum in this case occurs for a shrinkage of* 0.97. *A quasi-minimum variance portfolio achieved this way is said to be "weak" in the accompanying code, since the covariance has been weakened.*

example with four assets

$$\Sigma = \begin{pmatrix} 1.09948514 & -1.02926114 & 0.22402055 & 0.10727343 \\ -1.02926114 & 2.54302628 & 1.05338531 & -0.12481515 \\ 0.22402055 & 1.05338531 & 1.79162765 & -0.78962956 \\ 0.10727343 & -0.12481515 & -0.78962956 & 0.86316527 \end{pmatrix}$$

The minimum variance portfolio is

$$w = \frac{\Sigma^{-1}\vec{1}}{\vec{1}^T\Sigma^{-1}\vec{1}} = \begin{pmatrix} -9.008 \\ -6.871 \\ 8.749 \\ 8.130 \end{pmatrix}$$

and one can imagine that this might not be the best performing out of sample. Remarkably, if we take Σ as above and multiply off-diagonal by $\xi = 0.97$ (or divide diag by $1/\xi$) then the minimum variance portfolio is very different indeed:

$$w = \begin{pmatrix} 0.0674 \\ -0.0068 \\ 0.3658 \\ 0.5735 \end{pmatrix}$$

Now in this particular example the choice $\xi = 0.97$ doesn't come from thin air but from Figure 1. The value is the exact amount of shrinkage required to minimize the implied long-only portfolio. Usually this leaves a mall negative weight arising from optimization tolerance, which can be redistributed to the other three assets if a long=only portfolio is desired.

(Initially I experimented with more targeted weakening of the covariance matrix, such as reducing the cov entries only for those assets whose weights were initially negative. There's more to experiment with here.)

11. Weak covariance shrinkage. (work in progress)

If weak portfolios perform well in practice, which they seem to do, then this obviously suggests a method of shrinking covariance estimates. The question arises as to how effective this might be relative to, say, Ledoit-Wolf, Oracle Approximating Shrinkage, or something else.

12. Unifying parity, minimum vol, inverse variance and min variance?. (work in progress)

As noted in the construction of top-down allocation schemes, there are two uses of portfolios. One is the creation of sub-portfolio weights. The other is the computation of a metric, say portfolio variance that is used for inter-group allocation.

Can several of these appraoches be unified? It may help to work with forces rather than energy. First risk parity:

$$w_i(\Sigma w)_i = w_j(\Sigma w)_j \tag{12.1}$$

for all i, j is motivated by the notion that the weighted derivative of portfolio variance

$$w_i \frac{\partial \nu}{\partial w_i}$$

should be the same for each asset. On the other hand the minimum variance condition leads to

$$(\Sigma w)_i = (\Sigma w)_j \tag{12.2}$$

since both are equal to the Lagrange multiplier. (Notice that $w = \lambda \Sigma^{-1} \overrightarrow{1}$ gives rise to 2.2 and so Σw is a constant vector). A more basic portfolio also used in hierarchical risk parity allocates using inverse variance only:

$$\Sigma_{ii} w_i = \Sigma_{jj} w_j \tag{12.3}$$

and one might also consider inverse variance

$$\sqrt{\Sigma_{ii}} w_i = \sqrt{\Sigma_{jj}} w_j \tag{12.4}$$

which is often used as a first guess for risk parity portfolio weights. One can unify 12.1, 12.2, 12.3 and 12.4 by writing the balance equation

$$w_i^{\phi_1} \left((\Sigma - \xi \Sigma_o) w^{\phi_2} \right)_i = w_j^{\phi_1} \left((\Sigma - \xi \Sigma_o) w^{\phi_2} \right)_j \tag{12.5}$$

for all i, j where Σ_o denotes the off-diagonal matrix:

$$(\Sigma_o)_{ij} = \begin{cases} 0, i = j \\ \xi \Sigma_{ij}, i \neq j \end{cases}$$

is the covariance matrix with shrunk off-diagonal entries, parameters $\phi_1, \phi_2, \xi \in (0,1)$. Then we have the following examples

ϕ_1	ξ	ϕ_2	Interpretation
1	0	1	Risk parity
0	0	1	Minimum variance
0	ξ^*	1	Weak
0	1	1	Inverse variance
0	1	2	Inverse standard deviation

and this table also suggests some new possibilities. Here ξ^* denotes the shrinkage parameter defined in Section 10. It isn't necessarily trivial to solve 12.5 numerically, although cyclical coordinate descent, as used in parity, might rapidly converge.

13. Empirical findings: covariance estimation. Table 4 provides a rough indication of efficacy for a number of different covariance estimation procedures when weekly equity returns are estimated using only the last twenty weeks. As can been seen the running empirical covariance is a poor estimator compared to many other shrinkage approaches, including Oracle Approximating Shrinkage, Ledoit Wolf, partial moments and methods employing Huber pseudo-means. It is rather striking that the use of "weak" shrinkage occupies the podium, and improves existing methods.

A similar table is 5 where the covariance matrices are less rank-challenged, since at least 150 past data points are available to fix a 50×50 matrix (although whether the methods choose to use the entire history is up to them). Ledoit

REFERENCES

Elo	Approach	CPU
1691	Weak expon weighted Ledoit Wolf r=0.02	N/A
1681	Weak expon weighted Ledoit Wolf r=0.01	N/A
1670	ewa Ledoit Wolf r=0.02	N/A
1669	Buffered Oracle Approximating window=100	N/A
1659	Weak expon weighted partial moments empirical r=0.05 window=50 target=0	N/A
1652	Weak expon weighted Ledoit Wolf r=0.05	N/A
1635	Buffered Oracle Approximating window=300	N/A
1614	ewa Ledoit Wolf r=0.01	N/A
1599	ewa partial moments empirical r=0.02 window=100 target=0	N/A
1585	Buffered Ledoit Wolf window=300	N/A
1582	Buffered Ledoit Wolf window=200	N/A
1581	Buffered Oracle Approximating window=200	N/A
1573	ewa partial moments empirical r=0.02 window=100	N/A
1568	ewa partial moments empirical r=0.01 window=200	N/A
1563	ewa partial moments empirical r=0.02 window=50	N/A
1562	Weak expon weighted partial moments empirical r=0.02 window=100 target=0	N/A
1560	ewa partial moments empirical r=0.01 window=200 target=0	N/A
1560	Buffered Ledoit Wolf window=100	N/A
1559	ewa partial moments empirical r=0.05 window=25 target=0	N/A
1557	ewa partial moments empirical r=0.02 window=50 target=0	N/A
1551	ewa partial moments empirical r=0.01 window=100 target=0	N/A
1549	ewa partial moments empirical r=0.05 window=25	N/A
1549	ewa partial moments empirical r=0.01 window=100	N/A
1544	ewa partial moments empirical r=0.05 window=50	N/A
1529	ewa partial moments empirical r=0.05 window=50 target=0	N/A
1506	Expon weighted Lee Zhong r=0.005 r l=0.01 window=100	N/A
1503	Weak expon weighted empirical r=0.02	N/A
1489	Weak expon weighted empirical r=0.05	N/A
1475	Buffered Huber a=0.5b=2 window=200	N/A
1468	Buffered Huber a=0.5b=2 window=50	N/A
1465	Buffered Graphical lasso w/ cv window=200 target=0	N/A
1464	Buffered Huber a=0.5b=2 window=100	N/A
1457	Expon weighted Lee Zhong r=0.010 r l=0.01 window=100	N/A
1457	Weak expon weighted empirical r=0.01	N/A
1456	Expon weighted Lee Zhong r=0.005 r l=0.02 window=50	N/A
1455	ewa empirical r=0.01	N/A
1455	Buffered Huber a=1b=5 window=200	N/A
1454	Buffered empirical window=100	N/A
1453	ewa empirical r=0.10	N/A
1450	Buffered Graphical lasso window=100	N/A
1444	Buffered Graphical lasso w/ cv window=100	N/A
1442	ewa empirical r=0.02	N/A
1440	Buffered Huber a=1b=2 window=100	N/A
1425	Buffered Graphical lasso w/ cv window=100 target=0	N/A
1341	run empirical	N/A

TABLE 4

Performance on a rank-deficient covariance test. Equity 5 day returns (asset count 25) using historical data of length only 20. Elo ratings based on likelihood battles.

Elo	Approach	CPU
2123	Weak expon weighted partial moments empirical r=0.02 window=100 target=0	N/A
2117	Weak expon weighted Ledoit Wolf r=0.01	N/A
2094	ewa Ledoit Wolf r=0.01	N/A
2026	Expon weighted Lee Zhong r=0.010 r l=0.01 window=100	N/A
2003	Weak expon weighted partial moments empirical r=0.05 window=50 target=0	N/A
1982	Weak expon weighted empirical r=0.01	N/A
1920	Expon weighted Lee Zhong r=0.005 r l=0.02 window=50	N/A
1895	Buffered Graphical lasso w/ cv window=100 target=0	N/A
1891	Buffered Ledoit Wolf window=200	N/A
1882	Buffered Graphical lasso w/ cv window=300	N/A
1879	Buffered Graphical lasso w/ cv window=100	N/A
1874	Buffered Graphical lasso w/ cv window=200 target=0	N/A
1865	Buffered Ledoit Wolf window=300	N/A
1819	Buffered Graphical lasso window=100	N/A
1776	Buffered Ledoit Wolf window=100	N/A
1767	Buffered Graphical lasso w/ cv window=200	N/A
1752	Buffered Oracle Approximating window=100	N/A
1752	Buffered Graphical lasso window=200	N/A
1752	Buffered Oracle Approximating window=200	N/A
1749	Buffered Oracle Approximating window=300	N/A
1726	Buffered Ledoit Wolf window=300	N/A
1724	Weak expon weighted Ledoit Wolf r=0.02	N/A
1723	Weak expon weighted Ledoit Wolf r=0.05	N/A
1714	ewa Ledoit Wolf r=0.02	N/A
1683	Buffered Ledoit Wolf window=200	N/A
1630	Buffered Ledoit Wolf window=100	N/A
1581	ewa partial moments empirical r=0.02 window=100	N/A
1567	ewa partial moments empirical r=0.02 window=50	N/A
1563	ewa partial moments empirical r=0.02 window=100 target=0	N/A
1549	ewa partial moments empirical r=0.02 window=50 target=0	N/A
1543	ewa partial moments empirical r=0.05 window=25 target=0	N/A
1518	ewa partial moments empirical r=0.05 window=25	N/A
1506	Expon weighted Lee Zhong r=0.005 r l=0.01 window=100	N/A
1490	ewa partial moments empirical r=0.01 window=100	N/A
1459	ewa partial moments empirical r=0.01 window=200 target=0	N/A
1419	ewa empirical r=0.01	N/A
1373	run empirical	N/A
1340	ewa empirical r=0.02	N/A
1312	Buffered Huber a=0.5b=2 window=200	N/A
1305	Buffered Huber a=1b=5 window=200	N/A
1268	Buffered Huber a=1b=2 window=200	N/A
1165	Buffered empirical window=100	N/A
1148	Buffered Huber a=1b=5 window=100	N/A
1119	ewa partial moments empirical r=0.05 window=50 target=0	N/A
1105	Buffered Huber a=0.5b=2 window=100	N/A
1095	ewa partial moments empirical r=0.05 window=50	N/A
1044	ewa empirical r=0.05	N/A
1042	Buffered min-cov-det window=100	N/A
999	Weak expon weighted empirical r=0.05	N/A
996	Weak expon weighted empirical r=0.02	N/A
961	Buffered Huber a=1b=5 window=50	N/A
894	Buffered Huber a=1b=2 window=100	N/A
822	Buffered empirical window=50	N/A

TABLE 5

Cov likelihood for stocks 5 days with asset count 50 using historical data length 150.