SCHUR COMPLEMENTARY PORTFOLIOS

Top-down portfolio allocation is popular but lacks theoretical motivation. We exhibit a new family of recursive hierarchical methodologies where a reconciliation with optimization in the style of [Markowitz, 1952] is sometimes possible. This suggests a continuum where machine learning approaches such as [De Prado, 2016] lie at one extreme and optimization the other. We provide evidence that betwixt the two, superior out of sample performance can be attained.

Author's note. This is a provisional draft with literature review and final empirical results pending.

1. Seriate, divide and conquer. In a departure from optimization based portfolio construction ([Markowitz, 1952] and much that has followed), [De Prado, 2016] breathed new life into "top-down" capital allocation, specifically to the collection of methods where we begin by splitting our allocation between mutually exclusive collections of assets, and then proceed to sub-divide capital further within each one.

In that paper an idea that is retrospectively obvious, which is to say very good, is employed. One orders the assets in such a way that those most closely related financially (judged by price co-movement or another method, see [López de Prado, 2019] for an example) are nearest to each other in the sequence. That way, when we bisect into two groups there will be less inter-group information discarded. Alternatively put, the reordering reduces the size of the block off-diagonal terms in the covariance matrix - those which might as well be set to zero after the reordering since they will not enter subsequent calculations.

In this procedure one typically uses a defensive means of allocating capital between the two groups, which is to say one not leaning too heavily on off-diagonal elements in the (sub) covariance matrices, if at all. This combats high dimensional difficulties - though once we reach lower dimensions after a sufficient number of bisection steps, the procedure can terminate with some "conventional" portfolio optimization, if so desired.

Some, including the author of [De Prado, 2016] we infer, would term this a "machine learning" method because some ways of reordering assets might be so characterized. Certainly that phrase is also in keeping with the notion that the set of methods considered should not be constrained to the set for which theoretical results are available assuming a known generative model for the data. Possibly though, one might even call this an "archaeological approach" because, switching to the statistical terminology *seriation* to describe the structure conscious permutation of assets, we note that one of the oldest uses is in stratigraphy - see [Liiv, 2010] for an historical review.

Terminology to one side the goal of this paper is an examination, and partial remediation, of the information loss in divide-and-conquer asset allocation, and the reader can decide for themselves whether this should be labeled a unification of "machine learning" and "classical" technique. For it is certainly plausible that methods circumventing the need to invert a high dimensional covariance matrix can be beneficial when the number of assets (or models) approaches the number of historical data points used to estimate the covariance matrix.

Though many modifications can be suggested, we remark that the seriation will typically be robust to covariance shrinkage. Perhaps a hint as to the out of sample performance of hierarchical risk parity lies in this ambivalence. Unfortunately a thorough and compelling explanation is beyond the scope of this paper, or to our knowledge those works citing [De Prado, 2016]. If we are to take the empirical performance at face value, this is uncomfortable. Since top-down allocation can feel somewhat arbitrary, and is clearly suboptimal in the limit of perfect covariance knowledge, there would appear to be an unsatisfying hole in our comprehension.

Viewed this way, [De Prado, 2016] is a worthy provocation. We have two reactionary objectives. The first is to present a new way of relating the seemingly disparate approaches, by means of matrix inversion identities. The second is to determine empirically if this unifying view helps us locate allocation schemes "in between" the hierarchical and optimization extremal points that have superior performance to both.

2. Literature. Recent papers considering hierarchical allocation include [Raffinot, 2016], [Raffinot, 2018] and [Molyboga, 2020]. The deviations from [De Prado, 2016] might be characterized by modifications in the division of assets (aligning with clusters, rather than bisection); modifications to the measure of fitness used to allocate (denoted ν later in this paper); choices of covariance estimation technique; and changes to the terminal choice of portfolio. For instance [Molyboga, 2020] advocates Ledoit-Wolf shrinkage and the use of volatility rather than inverse portfolio variance.

This literature does not, however, contemplate augmentation of the sub matrices that are passed down from one step in the recursion to the next, and are therefore largely orthogonal to the novelty of this paper, as explained in Section 3. The same can be said of [Rane et al., 2022] or also [Kaczmarek and Perez, 2022] where, in passing, the authors push back on the empirical superiority of hierarchical risk parity over optimization.

The seriation step can be sensitive to choices of similarity. This is explored by [Lohre et al., 2020] although in other respects this work follows the established pattern. So too [Sen et al., 2021] and [Nourahmadi and Sadeqi, 2021] who appear to agree with [De Prado, 2016]'s sentiment regarding out of sample performance. It should be clear to the reader that there are many ways of performing clustering and partitioning of a portfolio, estimating correlations or covariances, and measures of similarity.

For instance [Pfitzinger and Katzke, 2021] consider similarity of assets near the bisection boundaries when those cuts are made, and claim that remaining conscious of potentially unfortunate loss of information can improve hierarchical performance (a goal similar to our own, though by a very different method). The reader might also be inspired by a survey [Marti et al., 2021] of graphical and other methods, and there would seem to be no reason why some of those tools could be plugged directly into top-down portfolio construction following the recipe provided by [De Prado, 2016].

Finally, it probably goes without saying that if optimization is used in the final step of recursive allocation then much of the literature on robust portfolio optimization is also relevant, including choices of shrinkage, portfolio weight constraints, and so forth. See [Becker et al., 2015], for example.

The class of hierarchical allocation methods, as they are presently understood, is certainly worthy of more study. But we note that regardless of the choices made - some of which may minimize information loss - there will be many who suspect the top-down allocation is wasteful, or simply inelegant, because any partitioning and recursion will inevitably suppress or ignore the relationships that probably exist between many pairs of assets.

3. Contribution. We introduce what we call "Schur complementary" portfolio construction to ameliorate this one very apparent drawback of recursive portfolio construction, namely the discarding of off-block diagonal covariance entries. The name comes from the role played by Schur matrix complements. Those corrections appear in matrix inversion identities that inspire the approach.

The qualitative distinction to [De Prado, 2016] and papers citing the same lies in the fact that (sub) covariance matrices used for inter and intra-group allocation in our new approach are not merely restrictions of the global covariance matrix. Instead they are augmented using some part of their mutual Schur complements. From a more probabilistic perspective, this might be viewed as an attempt to consider *conditional* rather than unconditional (sub) covariances, or some compromise between the two.

These modifications to the sub-covariance matrices we propose are not arbitrary but inspired by a desire to reconcile top-down allocation with "bottom-up" portfolio construction, where the latter refers to a single global optimization step. Seeking tractability, we consider unit minimum variance portfolios and demonstrate an equivalence. Though portfolio variance minimization may not be everyone's preference, other approaches can try the same matrix augmentations and, as with [De Prado, 2016] attempt to justify them by other means.

The unification presented here may enhance the ability to transfer the ideas to broader contexts, as also suggested in [De Prado, 2016]. For example there is an obvious isometry between minimum variance portfolio construction and minimizing the error of a convex combination of unbiased models in an ensemble. By explicitly relating this case to a top-down scheme we offer a new continuum for time-series model combinations [Cotton, 2021], mixtures of experts models, combinations of covariance estimators including those considered herein [Cotton, 2022b], or even the design of so-called micromanagers of models described in [Cotton, 2022a] who draw on a plurality of external predictions in order to create a superior ensemble.

4. Outline. Hierarchical risk parity seeks to address high dimensional challenges. Yet if we are to develop similar techniques on a firmer footing it is surely informative to consider low dimensional examples, and therein determine if information really needs to be discarded when recursive schemes are employed. It is our view that simple concrete examples are sufficient to motivate the new Schur recipe.

In Section 5 we review the basic mathematics of unit long-short portfolios (weights summing to unity) that minimize portfolio variance given perfect knowledge of covariance. In Section 6 we juxtapose these allocations against those achieved using more heuristic divide-and-conquer approaches, where the latter are viewed as "incorrect" notwithstanding the possibility of their empirical utility.

In Section 7 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.

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.

This is given in the same spirit as the original paper [De Prado, 2016], insofar as we are largely ambivalent as to the additional choices that need to be made, and invite the reader to experiment given code we have provided. However, we shall mention a new and quite simple portfolio variance estimator that in our studies has performed reliably in this context.

Section 8 and Section 9 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 role 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 in high dimensions. In Section 11 a comparison is performed using several decades of stock market return data and a recommendation is made for a reasonable compromise between the two philosophies. We shall not, in these experiments, rely on a single portfolio such as a popular index to make our empirical point. That would likely amount to p-hacking. Instead we consider randomly chosen portfolios over a very long time period.

There is still some danger here, and so in addition we complement our retrospective analysis in a novel way, by providing a single entry in the independently audited M6 Financial Forecasting Competition. This is obviously a rather limited exercise extending over only one year, but it may convey some probabilistic information to the Bayesian reader![1]

Finally, we also provide extensive numerical results in keeping with the spirit of [De Prado, 2016], where empirical results using actual financial data were entirely eschewed. We are sympathetic to the idea that this style of experiment might be the more informative, notwithstanding that synthetic data carries its own dangers.

5. A bottom up (optimization) example.. 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 bottom-up method of allocating capital to those three assets treats it as a constrained minimization of portfolio variance:

$$w(\Sigma) = \arg min_w w^t \Sigma w \quad s.t. \quad w^T \overrightarrow{1} = 1 \tag{5.1}$$

where for avoidance of doubt

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

The solution is well known.

$$w\left(\Sigma, 1\right) = \frac{\Sigma^{-1}\overrightarrow{1}}{\overrightarrow{1}^T \Sigma^{-1}\overrightarrow{1}} \tag{5.2}$$

For brevity we'll write

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

[1]See https://microprediction.github.io/precise for current standings.

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

$$w(\Sigma, b) = \arg min_w w^t \Sigma w \quad s.t. \quad w^t b = 1 \tag{5.3}$$

can be reduced to 5.1 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 5.2 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 5.2 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 5.1 should result in $w = (1/3, 1/3, 1/3)$. Indeed the application of 5.2 seems like hard work but nonetheless it will be useful to have an explicit representation on hand in what follows. So we observe:

$$\Sigma^{-1} = \frac{1}{1 - \rho - 2\rho^2} \begin{pmatrix} 1+\rho & -\rho & -\rho \\ -\rho & 1+\rho & -\rho \\ -\rho & -\rho & 1+\rho \end{pmatrix} \tag{5.4}$$

and then, continuing to ignore symmetry, diligently compute

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

which in generality is the portfolio variance when 5.1 is satisfied. This can be seen from 5.2, by writing $w_1 = w(\Sigma, 1)$ for brevity and noting:

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

Finally,

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

as expected.

In practical settings there may be other constraints other than represented in 5.1 or 5.3 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 this example, 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 5.2 is elegant, its use may not be advisable.

6. Top down portfolio construction. Next 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. As noted different things have been tried but 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 5.2. We can view any expression of the form $Q^{-1}\overrightarrow{1}$ in terms of the unit minimum variance long-short portfolio $w(Q)$ and its portfolio variance $\nu(Q)$, viz:

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

This, in passing, is a financial interpretation of the solution of a system of linear equations we will generalize later. For now, we can use it to design top down schemes because 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 $D \mapsto w(D)$.

Now although this is derived using 5.2 the form taken by this expression *suggests* generalizations beyond the case where the portfolio construction method w is minimum variance. That is the thread we shall pull on in due course, even when $B \neq 0$.

Potential weaknesses with this scheme might relate to the assumption $B = C^T = 0$, obviously, and let us walk through the first example with three assets to sharpen this concern. 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 this is not a terribly good general fix, 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.

Parenthetically, top-down approaches might be derived in other ways, such as 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 5.2 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 7.2 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.

We also hasten to add that existing technique can sometimes address these issues. For instance a mere reordering (possibly using seriation) might suffice, and other more careful schemes for clustering assets might reduce or eliminate the misallocation. 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.

Here we adopt the standpoint that these examples, simple as they are, present sufficient impetus for designing top-down portfolios 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.

7. Inversion identity. With that motivation, we next consider a family of new top-down approaches based on the following standard matrix inversion identity.

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

(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. 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. It is helpful to view some special cases of 7.1.

7.1. 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

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

A unit sub-portfolio determined by covariance A^c has weights
(7.2)
$$w(A^c,1) = \frac{(A^c)^{-1}\overrightarrow{1}}{\overrightarrow{1}^T(A^c)^{-1}\overrightarrow{1}} = \frac{\phi(\rho)}{2}\frac{1}{\phi}\begin{pmatrix}1+\rho & -\rho\\ -\rho & 1+\rho\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1+\rho & -\rho\\ -\rho & 1+\rho\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}$$

and we'll be using that expression shortly. Similarly

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

Here D^c is one dimensional so of course

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

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 5.4, namely:

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

Checking this, it is evident that $(\Sigma_3^{-1})_{3,3}$ corresponds to $1/D^c$ as required by 7.1. Also

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

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 5.4. 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 5.4. All is as it should be and reassembling the blocks, the identify 5.4 manifests in the case Σ_3 as

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

The right hand side has broken symmetry but we can check that weights of the minimum variance portfolio 5.2 are equal:

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

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 7.2 and we can suggestively write:

$$\begin{aligned} w \propto \Sigma^{-1}\overrightarrow{1} \quad & \propto \quad \begin{pmatrix} 1 & -\frac{\rho}{1+\rho} & 0 \\ -\frac{\rho}{1+\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1-\rho \\ 1-\rho \\ \frac{1-\rho}{1+\rho} \end{pmatrix} \\ & \propto \quad \begin{pmatrix} \begin{pmatrix} 1+\rho & -\rho \\ -\rho & 1+\rho \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ 1 \end{pmatrix} \\ & = \quad \begin{pmatrix} 2\ w(A^c) \\ w(D^c) \end{pmatrix} \end{aligned} \tag{7.3}$$

Our point: although this is the bottom-up solution, and of course trivial by symmetry, the intermediate steps in the calculation we are highlighting can now be read in top-down fashion. To wit, we allocate to assets $\{1,2\}$ and $\{3\}$ in the ratio $2:1$ and then split between assets 1 and 2 using the minimum variance portfolio *determined by the Schur complement* A^c. That is the key observation. And 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.

7.2. 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 previously 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

$$\begin{aligned} w(\Sigma) \propto \Sigma^{-1}\overrightarrow{1} &= \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A^c)^{-1} & 0 \\ 0 & (D^c)^{-1} \end{pmatrix} \begin{pmatrix} 1 & -BD^{-1} \\ -CA^{-1} & 1 \end{pmatrix} \overrightarrow{1} \\ &= \begin{pmatrix} (A^c)^{-1} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2\rho^2(1-\rho)}{1+\rho} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ (D^C)^{-1} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \rho & 0 \\ \rho & 0 \end{pmatrix} \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{pmatrix} \end{aligned}$$

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 8. The alternative route multiplies out the one vectors and will be discussed in Section 9.

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

$$\Sigma^{-1}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} \propto \begin{pmatrix}(W_A A^c)^{-1}\begin{pmatrix}1\\1\end{pmatrix}\\(W_D D^c)^{-1}\begin{pmatrix}1\\1\end{pmatrix}\end{pmatrix} \tag{7.4}$$

Once again, we have reduced the bottom-up global allocation using optimization to an equivalent, top-down scheme.

8. Schur complementary portfolios, method I. Equation 7.4 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

$$\Sigma^{-1}\overrightarrow{1} \propto \begin{pmatrix}\frac{1}{\nu(\tilde{A})}w(\tilde{A})\\ \frac{1}{\nu(\tilde{D})}w(\tilde{D})\end{pmatrix} \tag{8.1}$$

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.

We repeat the remark that this identify, and thus the top-down reinterpretation, holds for the minimum variance portfolio only, and yet it *suggests* generalization. Actually, we advocate the generalization that replaces $1 - BD^{-1}$ with $1 - \gamma BD^{-1}$ 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.[2]

9. Schur complementary portfolios, method II. There is an alternative to the use of W_A and W_D matrices. To get there we first consider the relationship

[2]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.

between $Q^{-1}b$ and portfolio weights in a slightly more general minimum variance problem. In particular a solution to

$$w* = \arg\min_{w} w^T Q w \;\; s.t. w^T b = 1 \tag{9.1}$$

is

$$w(Q,b) = \frac{Q^{-1}b}{b^T Q^{-1} b} \tag{9.2}$$

and the portfolio variance is

$$\nu(Q,b) = w(Q,b)^T Q w(Q,b) = \frac{1}{b^T Q^{-1} b} \tag{9.3}$$

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

$$Q^{-1}b = \frac{1}{\nu(Q,b)} w(Q,b) \tag{9.4}$$

in terms of the portfolio optimization 9.2 yielding $w(Q,b)$ and the resulting portfolio variance $\nu(Q,b)$. We had previous observed:

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

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

$$Q^{-1}b \leftrightarrow \frac{1}{\nu(Q^{*b})} \frac{w(Q_{/b})}{b} \tag{9.5}$$

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

$$Q_{/b} := \frac{Q}{bb^T} \tag{9.6}$$

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

$$Q^{*b} := \left(Q^{-1} \cdot (bb^T)\right)^{-1} \tag{9.7}$$

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

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

Allocation	HRP	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. Here HRP is hierarchical risk partiy.

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

$$w \propto \begin{pmatrix} (A^c(\gamma = 1))^{-1} b_A(\lambda = 1) \\ (D^c(\gamma = 1))^{-1} b_D(\lambda = 1) \end{pmatrix} \tag{9.8}$$

and thus the following top-down portfolio family can be "read" using 9.5

$$w(\Sigma; \lambda, \gamma) \propto \begin{pmatrix} \frac{1}{\nu\left((A^c(\gamma))^{*b_A(\lambda)}\right)} w\left(A^c(\gamma)_{/b_A(\lambda)}\right) \\ \frac{1}{\nu\left((D^c(\gamma))^{*b_D(\lambda)}\right)} w\left(D^c(\gamma)_{/b_D(\lambda)}\right) \end{pmatrix} \tag{9.9}$$

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

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 8 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 8 the modified versions of A and D are symmetric. See also Table 1 for a comparison against risk-parity.

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.

In some cases the choice of γ might be suggested by numerical properties of the matrices created. Many of our experiments first consider a maximal choice of γ that is the largest possible $\gamma < 1$ subject to the constraint that A^c is positive definite.

The continuum established is hopefully clear from Table 3 and the definitions $A^c(\lambda) = A - \lambda B D^{-1} C$ and $D^c(\lambda) = D - \lambda C A^{-1} B$. Also $b_A(\gamma) := \overrightarrow{1} - \gamma B D^{-1} \overrightarrow{1}$ and $b_D(\gamma) := \overrightarrow{1} - \gamma A C^{-1} \overrightarrow{1}$. 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.

10. Weak portfolios. Despite our philosophical ambivalence towards choices of $w(Q)$ and $\nu(Q)$, at least as far as the main message of this paper is concerned, we do require a defensive measure of portfolio variance and we are reluctant to be limited by those considered in the literature. Instead we discovered a very simple and unpublished method of robustly estimating portfolio variance through back-off, that is motivated financially.

Thus we introduce here the idea of "weak" portfolios, as we term them, as a variety of adaptive ridge estimator. The basic idea is to discourage covariance estimates that imply short positions (we assume that the appearance of the same is a bad omen).

This might be best illustrated by a simple concrete example. Suppose

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

The minimum variance portfolio is

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

and one can imagine that this might not be the best performing out of sample! There is great sensitivity. To personify, the optimizer doesn't really have a strong conviction that this long-short trade is worthwhile, it would seem, for if we multiply off-diagonal entries by $\xi = 0.97$ 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 not the variance of the portfolio but the variance of a portfolio *after* short positions have been nullified and the corresponding positive mass redistributed to all assets. The variance is judged by the original Σ, not the shrunk matrix.[3]

This choice of shrinkage usually leaves only small short positions. Those seeking a long-only portfolio can be assured there is probably one quite nearby. The approach is different to [DeMiguel et al., 2009] where the norm of weights are constrained, but shares a similar goal.

11. Provisional empirical findings. At present results are mostly "live" but snapshots will be incorporated into this paper.

11.1. Likelihood. The "weak" portfolio method uses a specific choice of shrinkage for covariance matrices. So it is of interest to see how this compares to Ledoit-Wolf, Oracle Approximating Shrinkage, and other well known strategies. Again, for the most up to date results, the reader is referred to the covariance Elo ratings in the precise package [Cotton, 2022b]. But for now we note that the defensive shrinkage adopted in the "weak" approach seems to lead to better ex-post likelihood measures.

However 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.

[3]We also experimented with more targeted weakening of the covariance matrix, such as reducing the cov entries only for those assets whose weights were initially negative. But we reverted to this more straightforward scheme.

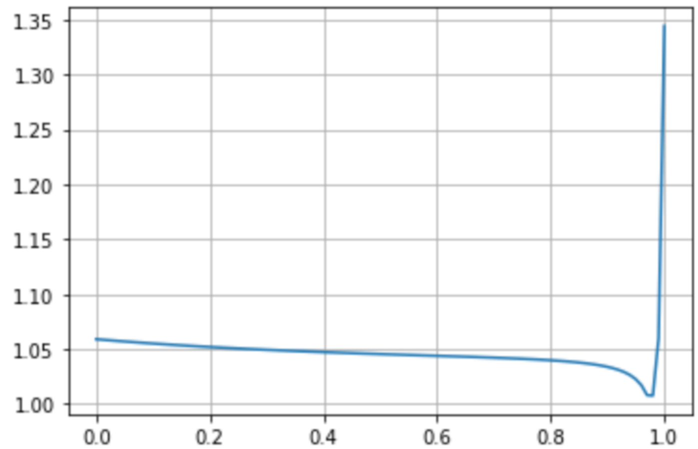

FIG. 1. *Portfolio variance of the long-only portfolio that is achieved by first multiplying off-diagonal entries and then redistributing the mass. The minimum in this case occurs for a shrinkage of* 0.97. *A quasi-minimum variance portfolio achieved this way is said to be "weak" in the accompanying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

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 as a technicality, whether the methods choose to use the entire history is entirely up to them).

11.2. Augmented hierarchical risk parity. To try to isolate the effect of (sub) covariance augmentation, comparisons were made between models using different values for γ. Some results are shown in Figures 2 and 3. A highly tentative finding is that adding small amounts of off-diagonal information seems to help, whereas the picture is more muddled when larger values for this parameter are chosen.

That finding is not terribly surprising. It can be read as a confirmation that [De Prado, 2016] presents a better guess ("about 90 percent right", speaking loosely) than dimension-challenged optimization at the other extreme (shrinkage of various kinds notwithstanding). Of course, we have also shown how to move ten percent of the way "back towards optimization" with results we think are likely to be even better out of sample.

REFERENCES

[Becker et al., 2015] Becker, F., Gürtler, M., and Hibbeln, M. (2015). Markowitz versus Michaud: portfolio optimization strategies reconsidered. *European Journal of Finance*, 21(4).

[Cotton, 2021] Cotton, P. (2021). Timemachines: A Python Package for Creating and Assessing Autonomous Time-Series Prediction Algorithms.

[Cotton, 2022a] Cotton, P. (2022a). *Microprediction: Building an Open AI Network*. MIT Press, Cambridge, MA.

[Cotton, 2022b] Cotton, P. (2022b). Precise: A Python Package for Streaming Covariance Prediction.

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

[DeMiguel et al., 2009] DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5).

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.

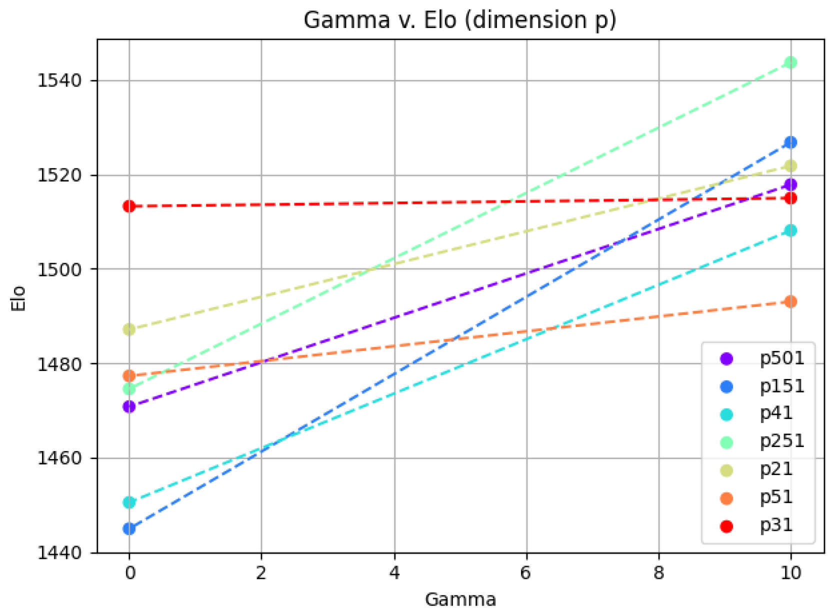

FIG. 2. *A test of whether including a small amount of off-diagonal information ($\gamma = 0.1$) helps hierarchical risk parity. A mix of models with $\gamma \in \{0, 0.1\}$ were used in repeated bake-offs using randomly selected stock universes dimension $p \in \{21, 31, 41, 51, 151, 501\}$. Elo ratings were computed based on the sorted ordering of Sharpe ratio. A warm-up period of 300 days was used, with prices provided every 3rd day. Then a subsequent period spanning 75 days was used for evaluation. All models used the same partial-moments covariance estimation, and the same "weak" method for both allocation and portfolio construction. So differences in seriation, inter-group allocation, and intra-group callibration arise only due to augmentation of the sub-covariance matrices. It is notable that independent of the dimensionality chosen, this augmentation appears to help, rather than harm.*

[Kaczmarek and Perez, 2022] Kaczmarek, T. and Perez, K. (2022). Building portfolios based on machine learning predictions. *Economic Research-Ekonomska Istrazivanja*, 35(1).

[Liiv, 2010] Liiv, I. (2010). Review seriation and matrix reordering methods: An historical overview.

[Lohre et al., 2020] Lohre, H., Rother, C., and Schäfer, K. A. (2020). Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-Asset Multi-Factor Allocations. *SSRN Electronic Journal*.

[López de Prado, 2019] López de Prado, M. (2019). Estimation of Theory-Implied Correlation Matrices. *SSRN Electronic Journal*.

[Markowitz, 1952] Markowitz, H. (1952). Portfolio Selection, Journal of Finance. *Markowitz HM—1952*.

[Marti et al., 2021] Marti, G., Nielsen, F., Bińkowski, M., and Donnat, P. (2021). A review of two decades of correlations, hierarchies, networks and clustering in financial markets. In *Signals and Communication Technology*.

[Molyboga, 2020] Molyboga, M. (2020). A Modified Hierarchical Risk Parity Framework for Portfolio Management. *Journal of Financial Data Science*, 2(3).

[Nourahmadi and Sadeqi, 2021] Nourahmadi, M. and Sadeqi, H. (2021). Hierarchical Risk Parity as an Alternative to Conventional Methods of Portfolio Optimization: (A Study of Tehran Stock Exchange). *Iranian Journal of Finance*, 5(4).

[Pfitzinger and Katzke, 2021] Pfitzinger, J. and Katzke, N. (2021). A Constrained Hierarchical Risk Parity Algorithm with Cluster-based Capital Allocation. Technical report.

[Raffinot, 2016] Raffinot, T. (2016). Hierarchical Clustering Based Asset Allocation. *SSRN Electronic Journal*.

[Raffinot, 2018] Raffinot, T. (2018). The Hierarchical Equal Risk Contribution Portfolio. *SSRN Electronic Journal*.

[Rane et al., 2022] Rane, C., Pai, S., Dani, M., and Dhage, S. (2022). Financial Portfolio Management and Optimization to Maximize Returns Using a Combination of HRP and Sentiment

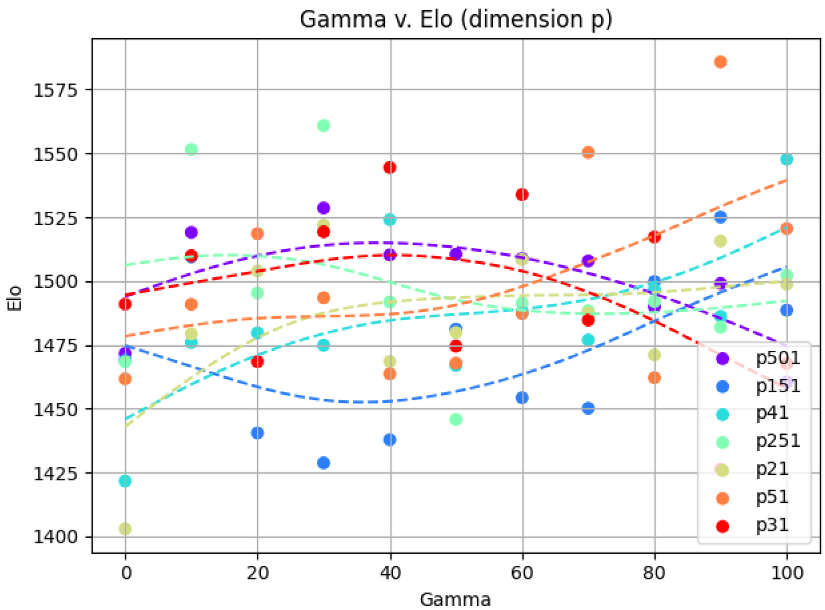

Fig. 3. *A test of whether adding larger amounts of off-diagonal information (i.e.* $\gamma \in (0.4, 1)$*) helps hierarchical risk parity. The Python smoothfit package has been used in an attempt to overlay some opinion on the noisy results. The picture is muddled, although amid the noise one is tempted to suspect diminishing returns on using higher values of* γ *in larger dimensions, but nonetheless using* $\gamma > 0$ *appears to to more to help than harm performance. We remark that in higher dimensions, there are more recursive steps taken and more opportunity for augmentation of sub-covariance matrices to occur.*

Analysis. In *Lecture Notes in Networks and Systems*, volume 237.

[Sen et al., 2021] Sen, J., Mehtab, S., Dutta, A., and Mondal, S. (2021). Hierarchical Risk Parity and Minimum Variance Portfolio Design on NIFTY 50 Stocks. In *2021 International Conference on Decision Aid Sciences and Application, DASA 2021.*