

Building Diversified Portfolios that Outperform Out of Sample

MARCOS LÓPEZ DE PRADO

MARCOS LÓPEZ DE PRADO

is a senior managing director at Guggenheim Partners in New York, NY, and a research fellow at the Lawrence Berkeley National Laboratory in Berkeley, CA.
lopezdeprado@lbl.gov

Portfolio construction is perhaps the most recurrent financial problem. On a daily basis, investment managers must build portfolios that incorporate their views and forecasts on risks and returns. Before Markowitz earned his Ph.D. in 1954, he left academia to work for the RAND Corporation, where he developed the Critical Line Algorithm (CLA), a quadratic optimization procedure specifically designed for inequality-constrained portfolio optimization problems. This algorithm is notable in that it guarantees finding the exact solution after a known number of iterations—and it ingeniously circumvents the Karush-Kuhn-Tucker conditions (Kuhn and Tucker [1952]). A description and open-source implementation of this algorithm can be found in Bailey and López de Prado [2013]. Surprisingly, most financial practitioners still seem unaware of CLA, as they often rely on generic-purpose quadratic programming methods that do not guarantee the correct solution or a stopping time.

Despite of the brilliance of Markowitz's theory, CLA solutions are somewhat unreliable because of a number of practical problems. A major caveat is that small deviations in the forecasted returns cause CLA to produce very different portfolios (Michaud [1998]). Given that returns can rarely be forecasted with sufficient accuracy, many authors have opted to drop them altogether and focus on the covari-

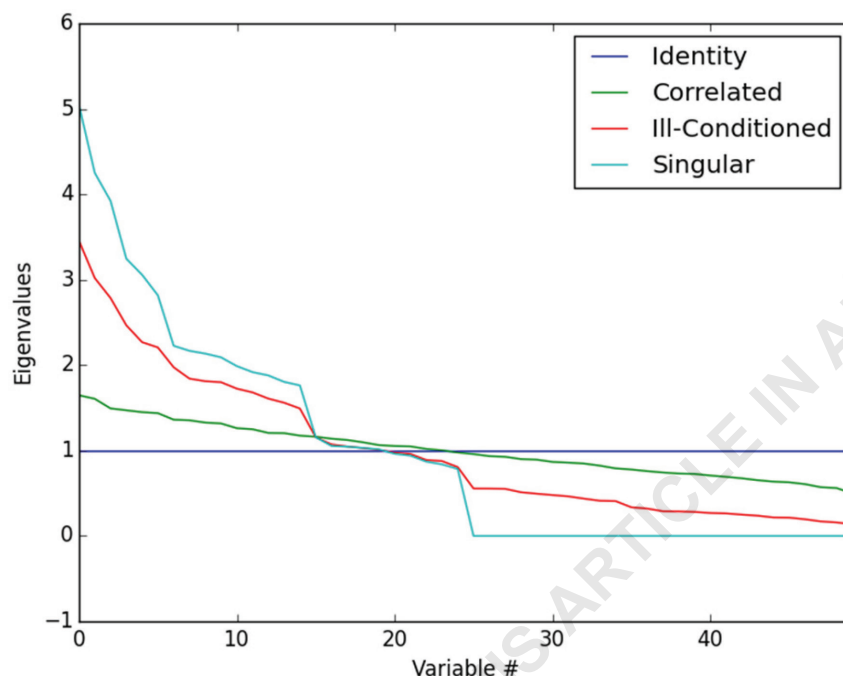
ance matrix. This has led to risk-based asset allocation approaches, of which “risk parity” is a prominent example (Jurczenko [2015]). Dropping the forecasts on returns improves the instability issues; however, it does not prevent them, because quadratic programming methods require the inversion of a positive-definite covariance matrix (all eigenvalues must be positive). This inversion is prone to large errors when the covariance matrix is numerically ill-conditioned—that is, it has a high condition number (Bailey and López de Prado [2012]).

MARKOWITZ'S CURSE

The condition number of a covariance, correlation (or normal, thus diagonalizable) matrix is the absolute value of the ratio between its maximal and minimal (by moduli) eigenvalues. Exhibit 1 plots the sorted eigenvalues of several correlation matrices, where the condition number is the ratio between the first and last values of each line. This number is lowest for a diagonal correlation matrix, which is its own inverse. As we add correlated (multicollinear) investments, the condition number grows. At some point, the condition number is so high that numerical errors make the inverse matrix too unstable: A small change on any entry will lead to a very different inverse. This is *Markowitz's*

EXHIBIT 1

Visualization of Markowitz's Curse



Notes: A diagonal correlation matrix has the lowest condition number. As we add correlated investments, the maximum eigenvalue is greater and the minimum eigenvalue is lower. The condition number rises quickly, leading to unstable inverse correlation matrices. At some point, the benefits of diversification are more than offset by estimation errors.

curse: The more correlated the investments, the greater the need for diversification—and yet the more likely we will receive unstable solutions. The benefits of diversification often are more than offset by estimation errors.

Increasing the size of the covariance matrix will only make matters worse, because each covariance coefficient is estimated with fewer degrees of freedom. In general, we need at least $\frac{1}{2}N(N+1)$ independent and identically distributed (IID) observations to estimate a covariance matrix of size N that is not singular. For example, estimating an invertible covariance matrix of size 50 requires at the very least five years of daily IID data. As most investors know, correlation structures do not remain invariant over such long periods by any reasonable confidence level. The severity of these challenges is epitomized by the fact that even naïve (equally weighted) portfolios have been shown to beat mean–variance and risk-based optimization out of sample (De Miguel, Garlappi, and Uppal [2009]).

FROM GEOMETRIC TO HIERARCHICAL RELATIONSHIPS

These instability concerns have received substantial attention in recent years, which has been carefully documented by Kolm, Tutuncu, and Fabozzi [2010]. Most alternatives attempt to achieve robustness by incorporating additional constraints (Clarke, De Silva, and Thorley [2002])—introducing Bayesian priors (Black and Litterman[1992]) or improving the numerical stability of the covariance matrix's inverse (Ledoit and Wolf [2003]).

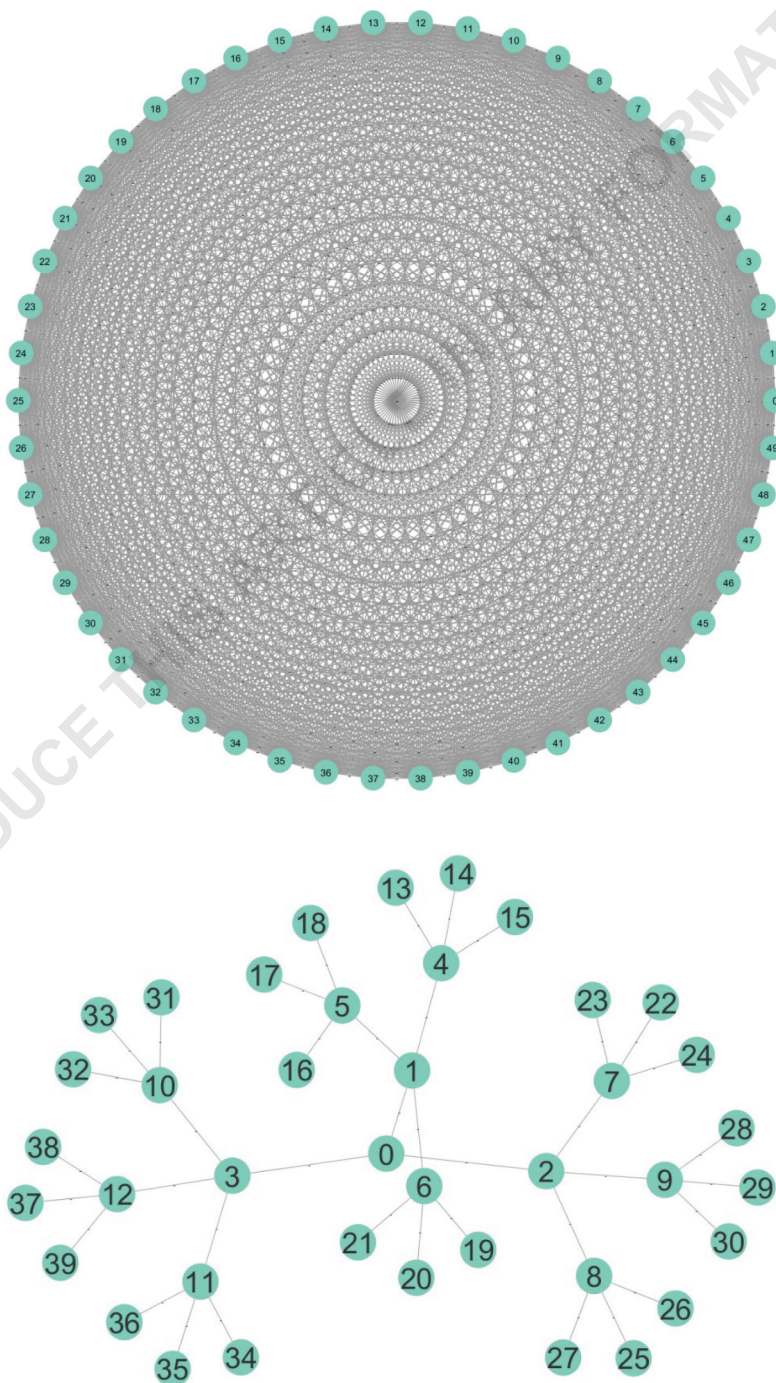
Although all the methods discussed so far were published in recent years, they are derived from (very) classical areas of mathematics: geometry, linear algebra, and calculus. A correlation matrix is a linear algebra object that measures the cosines of the angles between any two vectors in the vector space formed by the returns series (see Calkin and López de Prado [2014a, 2014b]). One reason for the instability of quadratic optimizers is that the vector space is modelled as a complete (fully connected) graph,

in which every node is a potential candidate to substitute for another. In algorithmic terms, inverting the matrix means evaluating the partial correlations across the complete graph. The top figure of Exhibit 2 visualizes the relationships implied by a covariance matrix of 50x50—that is, 50 nodes and 1225 edges. Small estimation errors are magnified, leading to incorrect solutions. Intuitively, it would be desirable to drop unnecessary edges.

Let's consider for a moment the practical implications of such topological structure. Suppose that an investor wishes to build a diversified portfolio of securities, including hundreds of stocks, bonds, hedge funds, real estate, private placements, etc. Some investments seem to be closer substitutes of one another, and other investments seem to be complementary to one another. For example, stocks could be grouped in terms of liquidity, size, industry, and region, in which stocks within a given group compete for allocations. In determining the allocation to a large publicly traded U.S. financial stock such as J.P. Morgan, we will consider adding or reducing the allocation to another large publicly traded U.S. bank such as Goldman Sachs, rather than a small community bank in Switzerland or a real estate holding in the Caribbean. And yet, in a correlation matrix, all investments are potential substitutes to each other; in other words, correlation matrices lack the notion of *hierarchy*.

The lack of hierarchical structure in a correlation matrix allows weights to vary freely in unintended ways, which is a root cause of CLA's instability. The bottom figure of Exhibit 2 visualizes a hierarchical structure known as a *tree*. A tree structure introduces two desirable features: 1) It has only $N - 1$ edges to connect N nodes, so the weights only rebalance among peers at various hierarchical levels; and 2) the weights are distributed top down, consistent with the way many asset managers build their portfolios (e.g., from asset class to sectors to individual securities). For these reasons, hierarchical structures are designed to give not only stable but also intuitive results.

EXHIBIT 2 The Complete-Graph (top) and the Tree-Graph (bottom) Structures



Notes: Correlation matrices can be represented as complete graphs, which lack the notion of hierarchy: Each investment is substitutable with another. In contrast, tree structures incorporate hierarchical relationships.

THE ALGORITHM

In this article, we study a new portfolio construction method that addresses CLA’s pitfalls using modern mathematics: graph theory and machine learning. This Hierarchical Risk Parity (HRP) method uses the information contained in the covariance matrix without requiring its inversion or positive definitiveness. In fact, HRP can compute a portfolio based on a singular covariance matrix—an impossible feat for quadratic optimizers. The algorithm operates in three stages:

1. *Tree clustering*: This stage combines investments into a hierarchical structure of clusters, so that allocations can flow downstream through a tree graph. In practical terms, think how MSCI classifies stocks at multiple levels (10 sectors, 24 industry groups, 67 industries, 156 subindustries) using a combination of quantitative and qualitative heuristics.

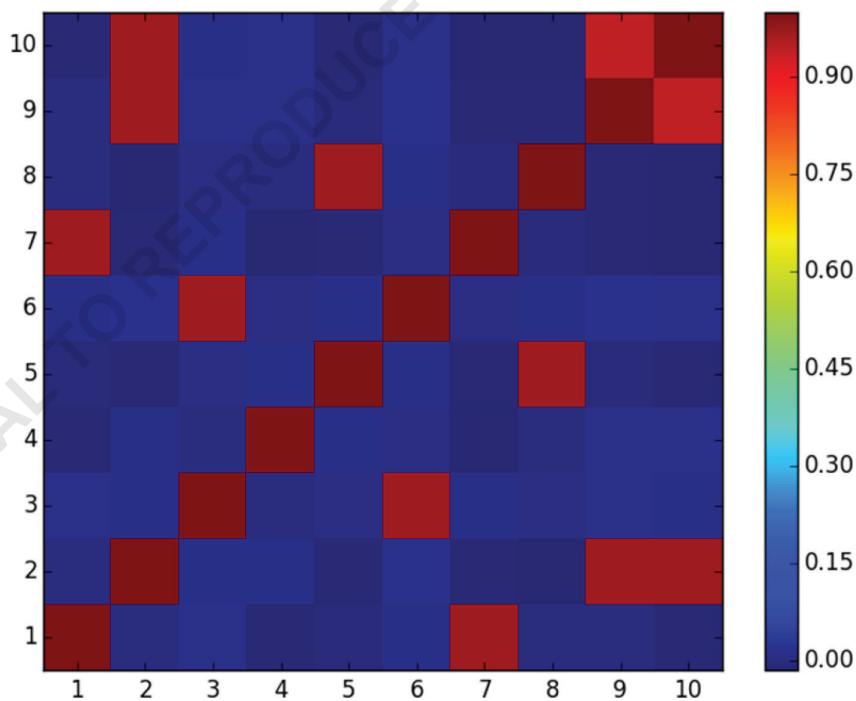
Tree clustering achieves the same objective in a mathematically rigorous way.

2. *Quasi-diagonalization*: This stage reorganizes the rows and columns of the covariance matrix, so that the largest values lie along the diagonal. This quasi-diagonalization of the covariance matrix (without requiring a change of basis) renders a useful property: Similar investments are placed together, and dissimilar investments are placed far apart (see Exhibits 4 and 5 for an example).
3. *Recursive bisection*: The inverse-variance allocation is optimal for a diagonal covariance matrix. We can take advantage of this fact by allocating funds top down through the tree structure, in which riskier clusters are given less funding.

The interested reader can find technical details and code in the online supplementary materials available at <http://ssrn.com/abstract=2708678>.

EXHIBIT 3

Heatmap of Original Covariance Matrix



Notes: This correlation matrix was computed on random series $X = \{X_i\}_{i=1,\dots,10}$ and drawn as follows. First, we drew five random vectors from a standard normal distribution, $\{X_j = z\}_{j=1,\dots,5}$. Second, we drew five random integer numbers from a uniform distribution, with replacement, $\mathfrak{D} = \{\mathfrak{D}_k\}_{k=1,\dots,5}$. Third, we computed $X_{5+k} = X_{\mathfrak{D}_k} + \frac{1}{4}z$, $\forall k = 1, \dots, 5$. This forced the last five columns to be partially correlated to some of the first five series.

A NUMERICAL EXAMPLE

We begin by simulating a matrix of observations X , of order (10000×10) . The correlation matrix is visualized in Exhibit 3 as a heatmap. Exhibit 4 displays the dendrogram of the resulting clusters (stage 1). Exhibit 5 shows the same correlation matrix, reorganized in blocks according to the identified clusters (stage 2). The online supplementary materials provide the code used to generate this numerical example.

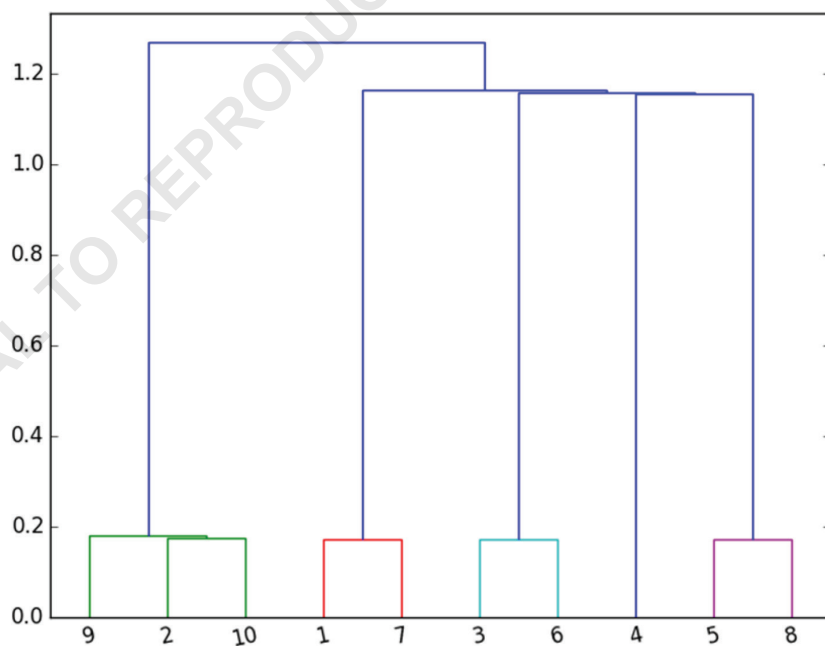
Using this random data, we compute HRP's allocations (stage 3), and compare them to the allocations from two competing methodologies: 1) quadratic optimization, as represented by CLA's minimum-variance portfolio (the only portfolio of the efficient frontier that does not depend on returns' means); and 2) traditional risk parity, exemplified by the Inverse-Variance Portfolio (IVP). (See Bailey and López de Prado [2013] for a comprehensive implementation of CLA, and Appendix A.2 of the online supplementary materials for a derivation of IVP.) We apply the standard constraints: $0 \leq w_i \leq 1$ (non-negativity), $\forall i = 1, \dots, N$, and $\sum_{i=1}^N w_i = 1$

(full investment). Incidentally, the condition number for the covariance matrix in this example is only 150.9324, which is not particularly high and, therefore, not unfavorable to CLA.

From the allocations in Exhibit 6, we can appreciate a few stylized features: First, CLA concentrates 92.66% of the allocation on the top five holdings, whereas HRP concentrates only 62.57%. Second, CLA assigns zero weight to three investments (without the $0 \leq w_i$ constraint, the allocation would have been negative). Third, HRP seems to find a compromise between CLA's concentrated solution and traditional risk parity's IVP allocation. The reader can use the code in Appendix A.3 of the online supplementary materials to verify that these findings generally hold for alternative random covariance matrices.

What drives CLA's extreme concentration is its goal of minimizing the portfolio's risk—and yet both portfolios have a very similar standard deviation ($\sigma_{HRP} = 0.4640$, $\sigma_{CLA} = 0.4486$). CLA discarded half of the investment universe in favor of a minor risk reduction. But the reality, of course, is that CLA's apparent diversi-

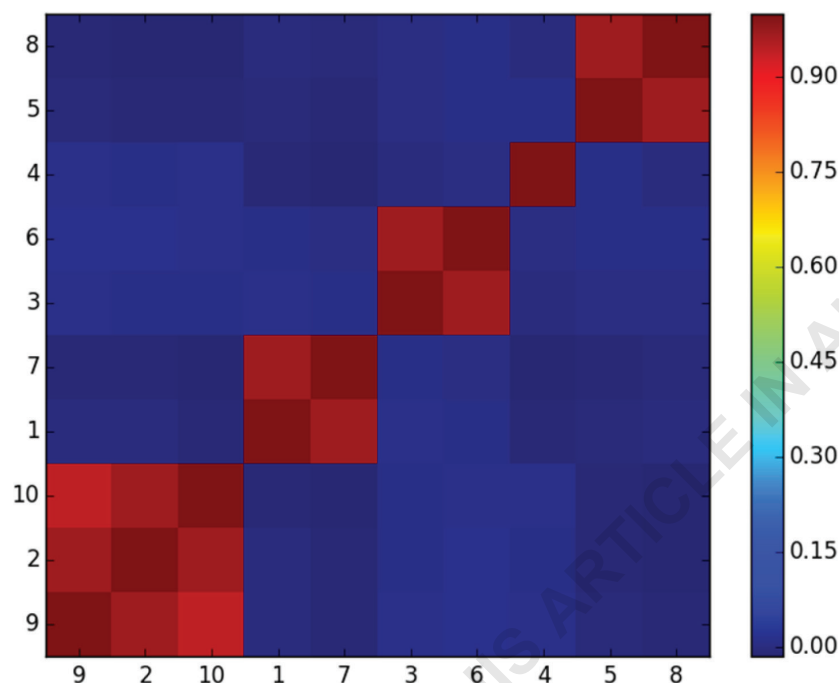
EXHIBIT 4 Dendrogram of Cluster Formation



Notes: The clustering procedure has correctly identified that series 9 and 10 were perturbations of series 2, and hence (9,2,10) are clustered together. Similarly, 7 is a perturbation of 1, 6 is a perturbation of 3, and 8 is a perturbation of 5. The only original item that was not perturbed is 4, and that is the one item for which the clustering algorithm found no similarity.

EXHIBIT 5

Clustered Covariance Matrix



Notes: Stage 2 quasi-diagonalizes the correlation matrix, in the sense that the largest values lie along the diagonal. However, unlike principal component analysis (PCA) or similar procedures, HRP does not require a change of basis. HRP solves the allocation problem robustly, while working with the original investments.

EXHIBIT 6

A Comparison of Three Allocations

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

Notes: This outcome is characteristic of the three methods studied: CLA concentrates weights on a few investments, hence becoming exposed to idiosyncratic shocks. IVP spreads weights evenly through all investments while ignoring the correlation structure, which makes it vulnerable to systemic shocks. HRP finds a compromise between diversifying across all investments and diversifying across clusters, which makes it more resilient against both types of shocks.

fiction is deceptive: Any distress situation affecting the top five allocations will have a much greater negative impact on CLA's portfolio than on HRP's.

OUT-OF-SAMPLE MONTE CARLO SIMULATIONS

In our numerical example, CLA's portfolio has lower risk than HRP's *in-sample* portfolio. However, the portfolio with minimum variance *in sample* is not necessarily the one with minimum variance *out of sample*. It would be all too easy for us to pick a particular historical dataset in which HRP outperforms CLA and IVP (for a discussion on overfitting and selection bias, see Bailey and López de Prado [2014]). Instead, in this section we evaluate, via Monte Carlo, the performance out of sample of HRP against CLA's minimum-variance and traditional risk parity's IVP allocations. This will also help us understand what features make a method preferable to the rest, regardless of anecdotal counter-examples.

First, we generate 10 series of random Gaussian returns (520 observations, equivalent to two years of daily history), with 0 mean and an arbitrary standard deviation of 10%. Real prices exhibit frequent jumps (Merton [1976]) and returns are not cross-sectionally independent, so we must add random shocks and a random correlation structure to our generated data. Second, we compute HRP, CLA, and IVP portfolios by looking back at 260 observations (a year of daily history). These portfolios are re-estimated and rebalanced every 22 observations (equivalent to a monthly frequency). Third, we compute the out-of-sample returns associated with those three portfolios. This procedure is repeated 10,000 times.

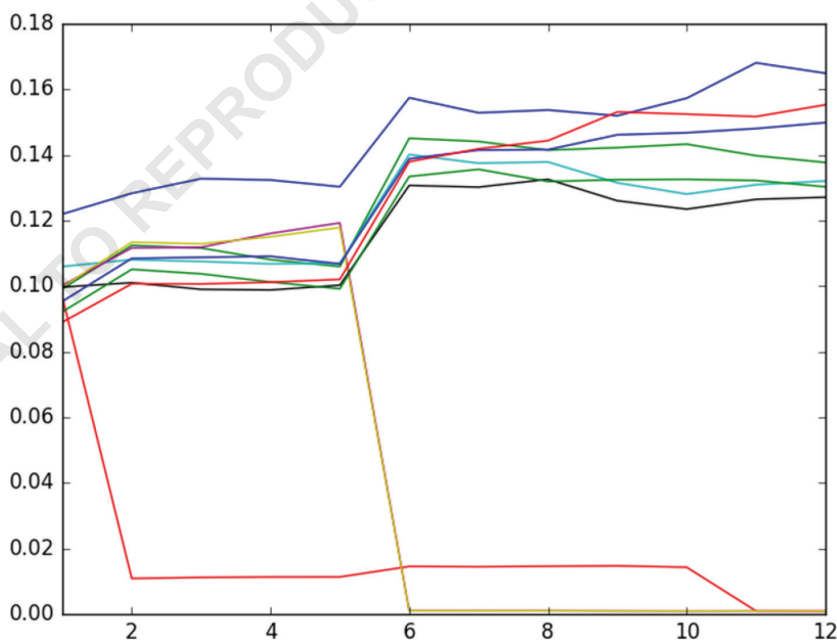
All mean portfolio returns out of sample are essentially zero, as expected. The critical difference comes from the variance of the out-of-sample portfolio returns: $\sigma_{CLA}^2 = 0.1157$, $\sigma_{IVP}^2 = 0.0928$, and $\sigma_{HRP}^2 = 0.0671$. Although CLA's goal is to deliver the lowest variance (that is, the objective of its optimization program), its performance happens to exhibit the highest variance out of sample, and a 72.47% greater variance than HRP's.

In other words, HRP would improve the out-of-sample Sharpe ratio of a CLA strategy by about 31.3%, a rather significant boost. Assuming that the covariance matrix is diagonal brings some stability to the IVP; however, its variance is still 38.24% greater than HRP's. This variance reduction out of sample is critically important to risk parity investors, given their use of substantial leverage. See Bailey et al. [2014] for a broader discussion of in-sample versus out-of-sample performance.

The mathematical proof for HRP's outperformance over Markowitz's CLA and traditional risk parity's IVP is somewhat involved and beyond the scope of this introductory article. Intuitively, we can understand the above empirical results as follows: Shocks affecting a specific investment penalize CLA's concentration. Shocks involving several correlated investments penalize IVP's ignorance of the correlation structure. HRP provides better protection against both common and idiosyncratic shocks by finding a compromise between diversification across all investments and diversification across clusters of investments at multiple hierarchical levels. Exhibit 7 plots the time series of allocations for the first of the 10,000 runs.

EXHIBIT 7

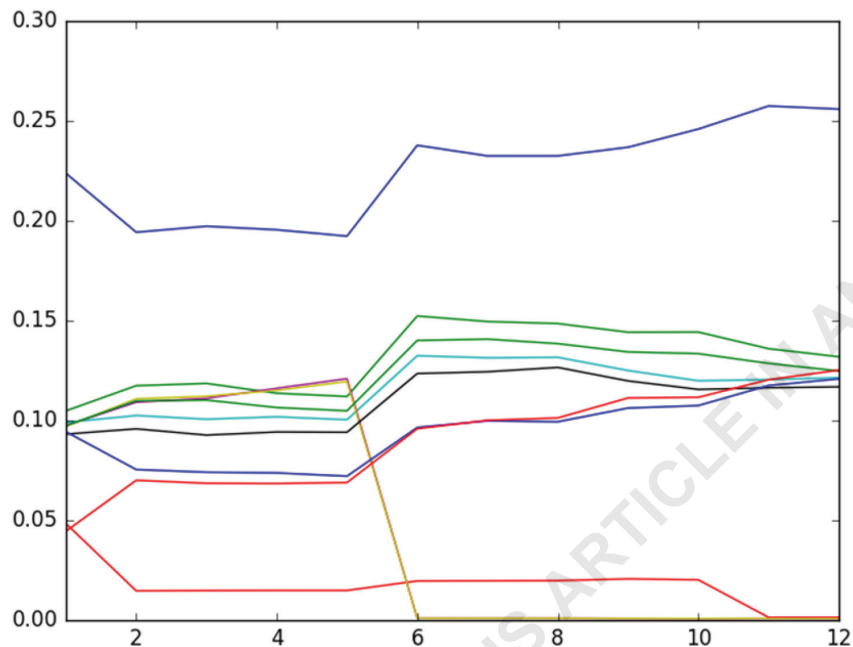
Panel A: Time Series of Allocations for IVP



Notes: Between the first and second rebalancing, one investment receives an idiosyncratic shock, which increases its variance. IVP's response is to reduce the allocation to that investment and spread that former exposure across all other investments. Between the fifth and sixth rebalancing, two investments are affected by a common shock; IVP's response is the same. As a result, allocations among the seven unaffected investments grow over time, regardless of their correlation.

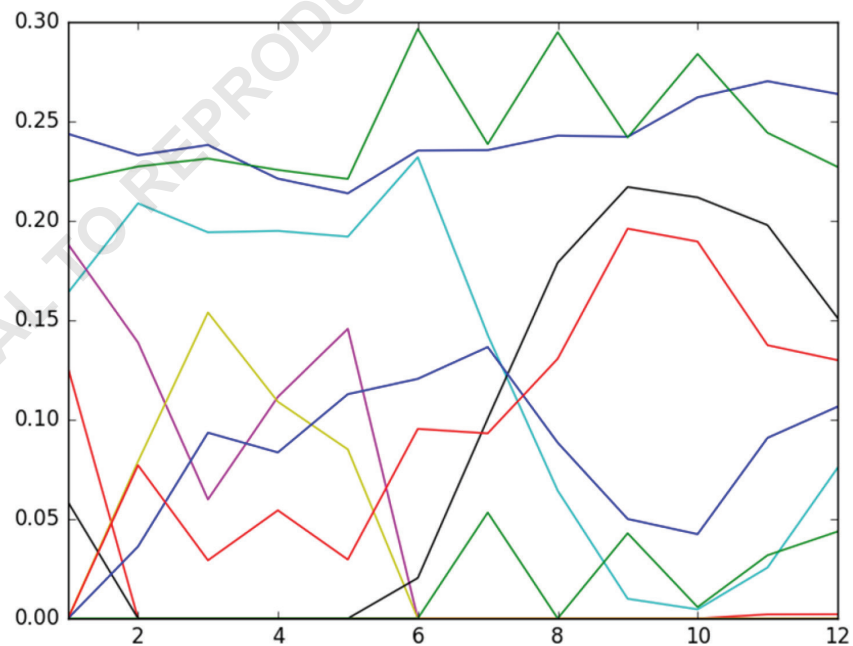
EXHIBIT 7 (Continued)

Panel B: Time Series of Allocations for HRP



Note: HRP's response to the idiosyncratic shock is to reduce the allocation to the affected investment, and use that reduced amount to increase the allocation to a correlated investment that was unaffected. As a response to the common shock, HRP reduces allocation to affected investments and increases allocation to uncorrelated ones (with lower variances).

Panel C: Time Series of Allocations for CLA



Note: CLA allocations respond erratically to idiosyncratic and common shocks. If we had taken into account rebalancing costs, CLA's performance would have been very negative.

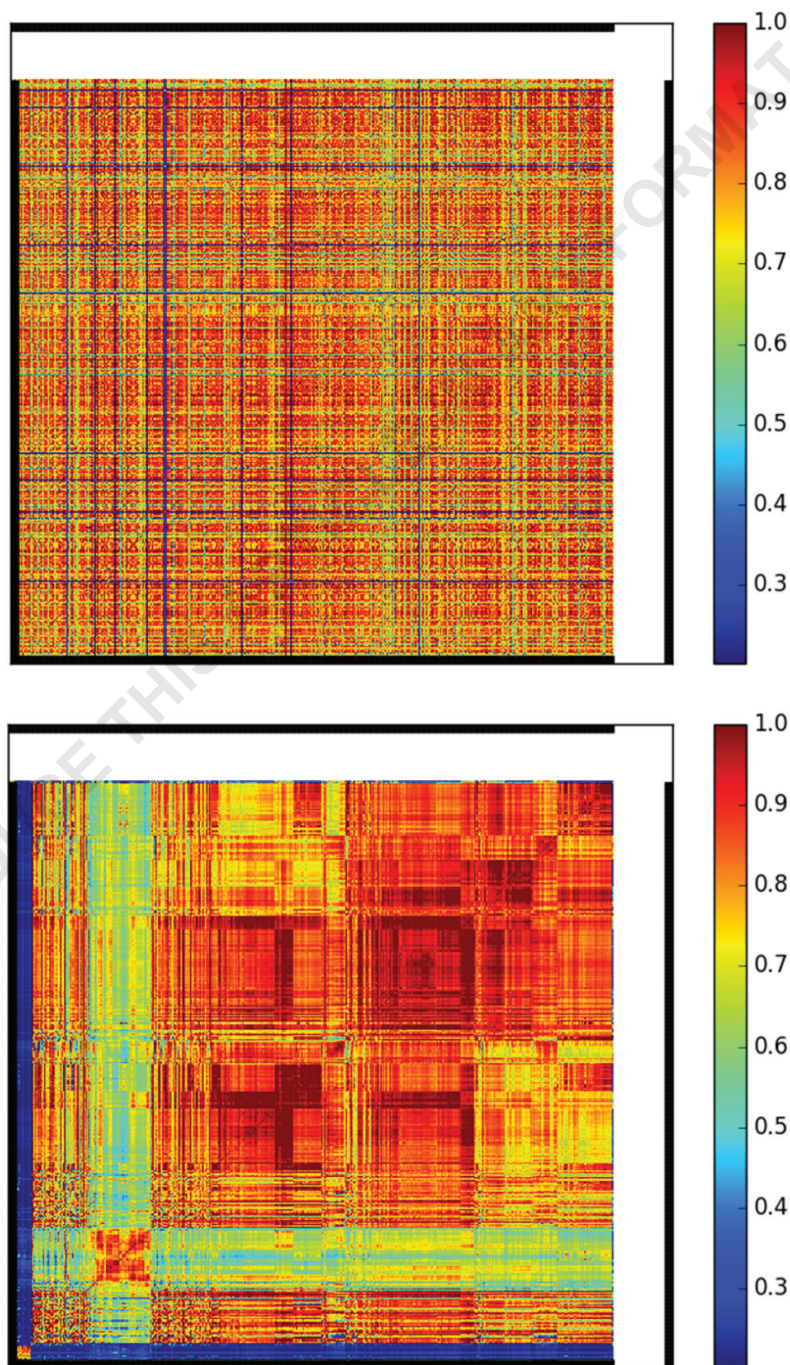
Appendix A.4 of the online supplementary materials provides the python code that implements the above study. The reader can experiment with different parameter configurations and reach similar conclusions. In particular, HRP's out-of-sample outperformance becomes even more substantial for larger investment universes, or when more shocks are added, or a stronger correlation structure is considered, or rebalancing costs are taken into account.

FURTHER RESEARCH

The methodology introduced in this article is flexible, scalable, and admits multiple variations of the same ideas. Using the code provided in the online supplementary materials, readers can research and evaluate what HRP configurations work best for their particular problem. For example, at stage 1, one can apply alternative definitions of $d_{i,j}$, $\tilde{d}_{i,j}$, and $\hat{d}_{i,w}$, or clustering algorithms; at stage 3, one can use different functions for \tilde{w}_m and α , or alternative allocation constraints. Instead of carrying out a recursive bisection, stage 3 could also split allocations top down using the clusters from stage 1.

In the future, we will show that it is relatively straightforward to incorporate forecasted returns and Black–Litterman-style views to this hierarchical approach. In fact, the inquisitive reader may have realized that, at its core, HRP is essentially a robust procedure to avoid matrix inversions, and the same ideas underlying HRP can be used to replace many econometric regression methods, notorious for their unstable outputs (such as vector autoregressive [VAR] or a vector error correction model [VECM]). Exhibit 8 displays a large correlation matrix of fixed income securities before and after clustering, with over 2.1 million entries. Traditional optimization or econometric methods fail to recognize the hierarchical structure of financial big data—where the numerical instabilities defeat the benefits of the analysis, resulting in unreliable and detrimental outcomes.

EXHIBIT 8 Correlation Matrix Before and After Clustering



Notes: The methodology described in this article can be applied to problems beyond optimization. For example, a PCA analysis of a large fixed income universe suffers the same drawbacks we described for CLA. Small-data techniques developed decades and centuries ago (factor models, regression analysis, econometrics) fail to recognize the hierarchical nature of financial big data.

CONCLUSIONS

Although mathematically correct, quadratic optimizers in general, and Markowitz's CLA in particular, are known to deliver generally unreliable solutions due to their instability, concentration, and underperformance. The root cause of these issues is that quadratic optimizers require the inversion of a covariance matrix. Markowitz's curse is that the more correlated investments are, the greater is the need for a diversified portfolio—and yet the greater are that portfolio's estimation errors.

In this article, we have exposed a major source of quadratic optimizers' instability: a matrix of size N is associated with a complete graph with $\frac{1}{2}N(N-1)$ edges. With so many edges connecting the nodes of the graph, weights are allowed to rebalance with complete freedom. This lack of hierarchical structure means that small estimation errors will lead to entirely different solutions. HRP replaces the covariance structure with a tree structure, which accomplishes three goals: a) unlike traditional risk parity methods, it fully utilizes the information contained in the covariance matrix, b) it recovers the stability of the weights, and c) the solution is intuitive by construction. The algorithm converges in deterministic logarithmic time.

HRP is robust, visual, and flexible, allowing the user to introduce constraints or manipulate the tree structure without compromising the algorithm's search. These properties are derived from the fact that HRP does not require covariance invertibility. Indeed, HRP can compute a portfolio on an ill-degenerated or even a singular covariance matrix—an impossible feat for quadratic optimizers.

This article focuses on a portfolio construction application, but the reader will find other practical uses for making decisions under uncertainty, particularly in the presence of a nearly-singular covariance matrix: capital allocation to portfolio managers, allocations across algorithmic strategies, bagging and boosting of machine-learning signals, forecasts from random forests, replacement to unstable econometric models (VAR, VECM), etc.

Of course, quadratic optimizers like CLA produce the minimum-variance portfolio in sample (that is, its objective function). Monte Carlo experiments show that HRP delivers lower out-of-sample variance than CLA or traditional risk parity methods (IVP). Since Bridgewater pioneered risk parity in the 1990s, some of the largest asset managers have launched funds that

follow this approach for combined assets in excess of \$500 billion. Given their extensive use of leverage, these funds should benefit from adopting a more stable risk parity allocation method—thus achieving superior risk-adjusted returns and lower rebalancing costs.

ENDNOTE

The statements made in this communication are strictly those of the author and do not represent the views of Guggenheim Partners or its affiliates. No investment advice or particular course of action is recommended. All rights reserved.

REFERENCES

- Bailey, D., and M. López de Prado. "Balanced Baskets: A New Approach to Trading and Hedging Risks." *Journal of Investment Strategies*, Vol. 1, No. 4 (2012), pp. 21-62.
- . "An Open-source Implementation of the Critical-line Algorithm for Portfolio Optimization." *Algorithms*, Vol. 6, No. 1 (2013), pp. 169-196.
- . "The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting and Non-Normality." *The Journal of Portfolio Management*, Vol. 40, No. 5 (2014), pp. 94-107.
- Bailey, D., J. Borwein, M. López de Prado, and J. Zhu. "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-Of-Sample Performance." *Notices of the American Mathematical Society*, Vol. 61, No. 5 (2014), pp. 458-471.
- Black, F., and R. Litterman. "Global Portfolio Optimization." *Financial Analysts Journal*, 48 (1992), pp. 28-43.
- Calkin, N., and M. López de Prado. "Stochastic Flow Diagrams." *Algorithmic Finance*, Vol. 3, No. 1 (2014a), pp. 21-42.
- . "The Topology of Macro Financial Flows: An Application of Stochastic Flow Diagrams." *Algorithmic Finance*, Vol. 3, No. 1 (2014b), pp. 43-85.
- Clarke, R., H. De Silva, and S. Thorley. "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal*, 58 (2002), pp. 48-66.
- De Miguel, V., L. Garlappi, and R. Uppal. "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22 (2009), pp. 1915-1953.

Jurczenko, E., ed. *Risk-Based and Factor Investing*. Oxford, UK: Elsevier Science, 2015.

Kolm, P., R. Tutuncu, and F. Fabozzi. "60 years of Portfolio Optimization." *European Journal of Operational Research*, Vol. 234, No. 2 (2010), pp. 356-371.

Kuhn, H.W., and A.W. Tucker. "Nonlinear programming." In *Proceedings of 2nd Berkeley Symposium*, pp. 481-492. Berkeley, CA: University of California Press, 1952.

Ledoit, O., and M. Wolf. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection." *Journal of Empirical Finance*, Vol. 10, No. 5 (2003), pp. 603-621.

Michaud, R. *Efficient Asset Allocation: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Boston, MA: Harvard Business School Press, 1998.

To order reprints of this article, please contact Dewey Palmieri at dpalmieri@ijournals.com or 212-224-3675.